The Dissertation Committee for Navin Varadarajan Certifies that this is the approved version of the following dissertation:

Engineering highly active and specific protease variants

Committee:

Brent Iverson, Supervisor

George Georgiou, Co-supervisor

Eric Anslyn

Dean Appling

Christian Whitman

Engineering highly active and specific protease variants

by

Navin Varadarajan, B.Sc.; M.S

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December 2006

Dedication

To my parents for their never ending love and support and to the rest of my family and friends.

Acknowledgements

I would like to thank my advisors Dr. Brent Iverson and Dr. George Georgiou for their constant support and encouragement throughout my graduate career.

I would also like to thank Mark Olsen, Daren Stephens, Edgardo Farinas, Jongsik Gam, Bum-Yeol Hwang, Marsha Demers and Sarah Rodriguez for working with me on the different projects.

It has been a great pleasure working with the members (both past and present) of the BIGG lab and the Iverson chemistry group.

Engineering highly active and specific protease variants

Publication No._____

Navin Varadarajan, PhD

The University of Texas at Austin, 2006

Supervisor: Brent Iverson

Co-supervisor: George Georgiou

The reprogramming of enzyme catalytic activity and selectivity is a central issue in protein biochemistry and biotechnology. Numerous structure-guided and directed evolution strategies have been employed in search of enzyme variants that exhibit high catalytic rates with poor or inactive substrates of the parental enzyme. As impressive as these successes have been, the engineering of enzymes that exhibit turnover rates and selectivities with new substrates comparable to their natural counterparts has proven quite a challenge, especially when considering those enzymes for which a genetic selection strategy is not possible.

By utilizing bacterial display and multi-parameter flow cytometry we have developed a novel methodology for emulating positive and negative selective pressure *in vitro* for the isolation of enzyme variants with reactivity for desired novel substrates, while simultaneously excluding those with reactivity towards undesired substrates.

In order to demonstrate the application of the high-throughput flow-cytometric for protease engineering; we sought to evolve a set of highly active OmpT variants that have P1 specificities altered systematically to recognize one amino-acid from each of the six classes of amino acids By screening error-prone and multiple residue saturation libraries we describe the systematic directed evolution of a set of proteases with altered recognition sites. A set of OmpT variants were engineered that can specifically cleave substrates having a hydrophobic, polar, aromatic and even acidic residue at the P1 and Arg at the P1'. In particular we note that the change in electrostatic specificity from a basic amino acid (Arg) to an acidic (Glu) is unprecedented.

After successfully changing P1 specificity, we then focused our attention on isolating OmpT mutants that recognize altered P1' specificities such as Ala and Val. Towards this end, we show the isolation of highly active OmpT variants that cleave Arg-Val and Glu-Ala sequences.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Enzyme engineering: Altering the substrate specificity of enzymes

1.1 Introduction

Enzymes are intricate biocatalysts whose specificity has been shaped by evolution in the context of their physiological substrates. These remarkable molecular machines have a tremendous advantage over traditional chemical catalysts not only because of high chemo- and stereoselectivity coupled with high rate accelerations. Advances in recombinant technology have made modification and production of enzymes relevant on the industrial biocatalysis scale (1),(2). One of the major obstacles in the application of enzymes as widespread catalysts is their specificity towards physiological substrates rather than industrially relevant ones. Enzyme engineering, either through rational design or combinatorial screening, can identify enzyme variants with modified properties like enhanced stability in organic solvents (3). One of the intriguing aspects of enzymes is their catalytic promiscuity, providing enzyme engineers an avenue to reprogram specificity (4). The ability to modify the specificity of enzymes can serve two functions; improving properties important for practical catalysis, while the other is increasing the

understanding of their mechanism and evolution from both a molecular and biological standpoint. For example, alteration of substrate specificity has been used to understand the evolution and mechanism of the *E. coli* chaperone GroEL/S (5), understand the evolution of microbial resistance to antibiotics (6), and discern accessible pathways in Darwinian evolution (7).

In the following sections, the enzymes have been classified by super-family according to the amount of research dedicated towards altering their substrate selectivity and/or mechanistic/substrate similarities. Wherever appropriate, the application of directed evolution towards understanding biological mechanisms has been highlighted.

1.2 Rational design of biocatalysts

Although the *de novo* design of biocatalysts is still an extremely challenging problem (8), there has been recent progress. For instance, Park *et al.* describe the design and evolution of new catalytic activity with an existing protein scaffold. The authors used SAIFE (simultaneous incorporation and adjustment of functional elements) that consists of insertions, deletions, and substitution of gene elements in conjunction with directed evolution to select for β-lacatamase activity in the structurally homologous glyoxalase II αβ/βα metallohydrolase (Figure 1.1) (9). Dwyer *et al.* on the other hand used structure-based computational design techniques that led to new catalytic activites in binding proteins (10). Using their methodology they were able to design a glycolytic enzyme analogous to triose phosphate isomerase based on the ribose-binding protein. Ueno *et al.*

demonstrated the design of metal cofactors activated by a protein-protein electron transfer system (11). Heme oxygenase catalyzes the conversion of heme to biliverdin, utilizing the electrons provided from a NADPH/cytochrome P_{450} reductase system (CPR). By designing a Fe(III) (Schiff-base) heme oxygenase as the electron acceptor, they were able to design a synthetic metal complex activated by a protein-protein electron transfer relay system (For a review on the rational design of heme enzymes please see (12)). Wei *et al.* used binary patterning of polar and non-polar residues (for a recent review please see (13)) to generate combinatorial *de novo* protein libraries. In order to test the feasibility of *de novo* designed folded proteins to yield catalytic enzymes, they tested the esterase activity of S-284, a four-helix bundle protein. S-284 displayed a 8700-fold rate enhancement for the hydrolysis of p-nitrophenyl esters over the uncatalyzed reaction (14). This rate enhancement is similar to that seen with hydrolytic catalytic antibodies (15).

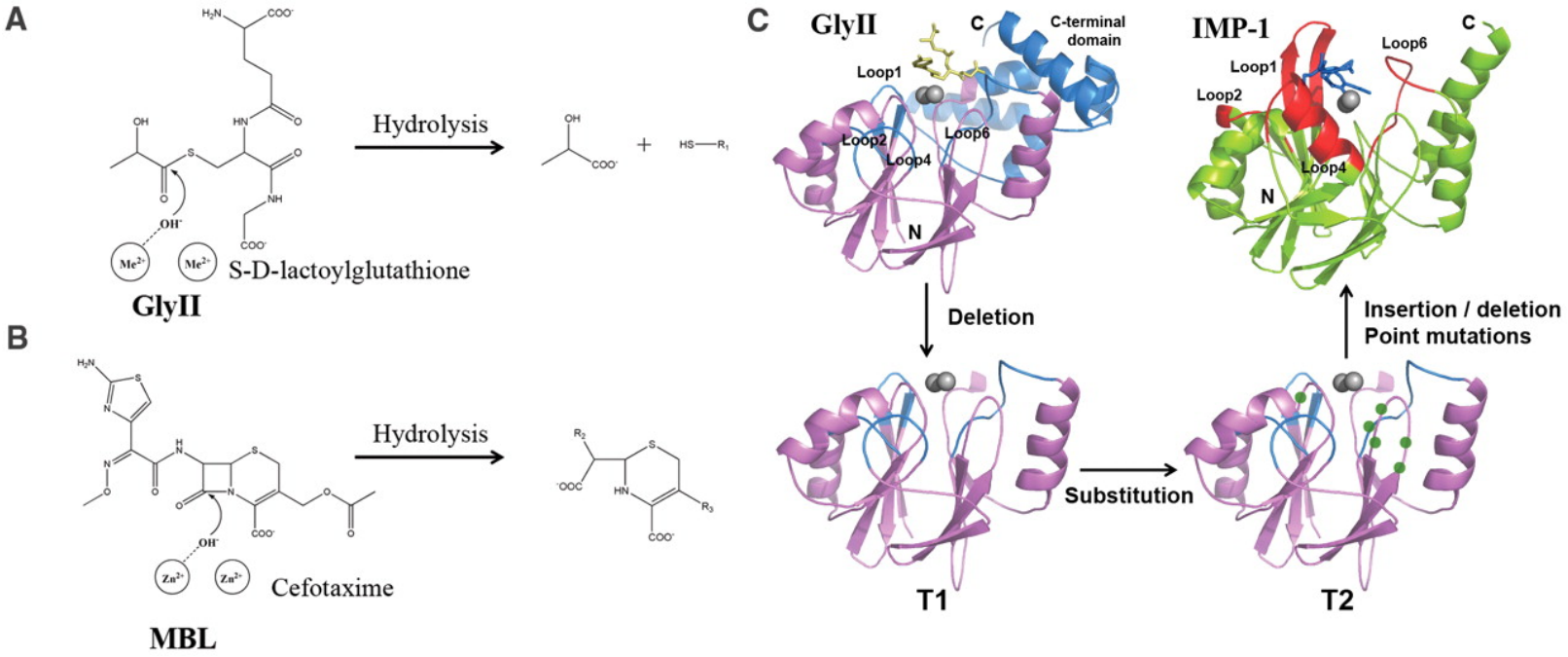


Figure 1.1: "Strategy for generation of MBL activity from GlyII. Catalytic mechanisms of (**A**) GlyII and (**B**) MBL. (**C**) Reconstruction of the active site of GlyII for generation of IMP-1 activity with the SIAFE process. As the first step, the *C*-terminal domain was deleted to produce template T1, and the catalytic elements (marked by green circles) were substituted to yield T2. Substrate-binding elements consisting of predesigned functional loops with different lengths and sequences were inserted, and random point mutations were induced for further adjustment. Functional loops constituting the active site of IMP-1 [Protein Data Bank (PDB) entry, 1DDK)] are shown in red, and the corresponding loops of GlyII (PDB entry, 1QH5) are in blue. The *C*-terminal domain to be deleted is shown in blue. Substrates for GlyII and IMP-1, S-(*N*-hydroxy-*N*-bromophenylcarbamoyl)glutathione and cefotaxime, are shown in yellow and blue, respectively." Figure and caption from Park *et al.* Science (2006) 311, 535. Reprinted with permission from AAAS.

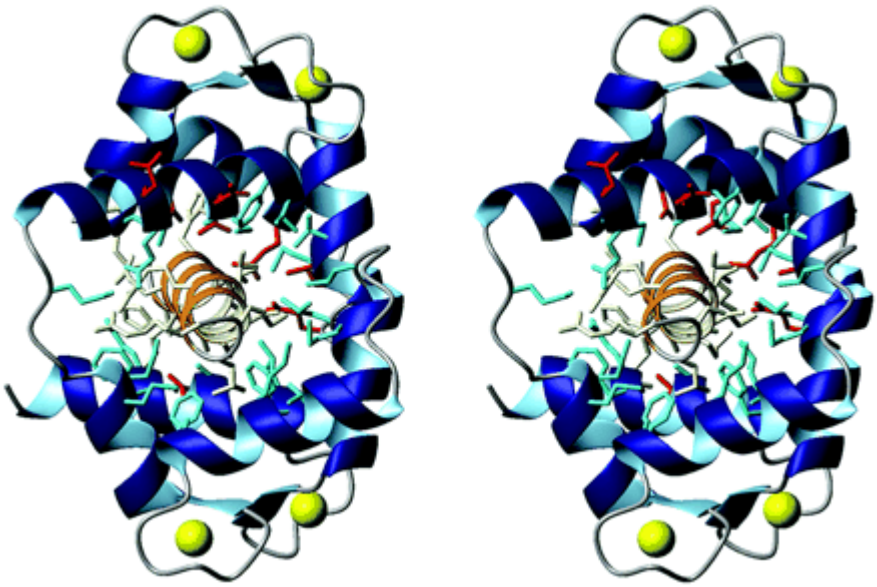
1.3 Rational redesign of substrate specificity.

The computational redesign of enzyme substrate specificity has been demonstrated for enzymes with well-characterized function and structures. Computational algorithms have been used not only to modify enzyme function but also to remodel protein-protein interfaces (16). Ashworth *et al.* have used a physically realistic atomic-level force field algorithm to redesign the cleavage specificity of the intron-encoded homing endonuclease I-MsoI (17). By using an *in silico* screen to disrupt the affinity of the protein towards the wild-type recognition sequence 5' -GCAGAA**C**GTCGTGAGAC**A**GTTCCG-3' and re-optimizing the amino acid cluster around the target recognition site 5' CAGAA**G**GTCGTGAGAC**C**GTTCCG 3' they designed and characterized a modified protein I-MsoI DES that had both lower experimental affinity and higher cleavage specificity for the target sequence. The experimental verification of the design algorithm highlighted its efficiency in the context of other engineered restriction endonucleases identified using directed evolution (discussed below) and has demonstrated the application of such algorithms as viable avenues for the re-design of substrate selectivity for enzymes with well-characterized structures.

The ability to commit a promiscuous enzyme to recognize/catalyze a specific substrate(s) has been demonstrated for at least two different enzyme systems. γ-Humulene synthase is a promiscuous sesquiterpene synthase that catalyzes the conversion of a single substrate, farnesyl diphosphate, to 52 different sesquiterpenes. Yoshikuni *et al.* identified residues that constitute the active-site contour and performed single-site saturation mutagenesis at

these sites and recombined them using a computational algorithm to yield a set of six γ-humulene synthases tailored to yield different terpenes (18). Again, Calmodulin (CaM) is a promiscuous enzyme that has evolved to bind to 14-30 amino acid stretches on a variety of target proteins. In order to tailor CaM to be specific for a smooth muscle myosin light chain kinase peptide (smMLCK, Figure 1.2), Shifman *et al*. selected for residues that simultaneously increased the affinity of the enzyme towards smMLCK, while decreasing the affinity towards alternative targets (19). Using an optimization algorithm that ensured preservation of fitness with the desired peptide while simultaneously exploring tolerated residues that decreased the affinity towards alternate targets, they designed a CaM variant that showed a 155-fold increase in specificity.

A



B

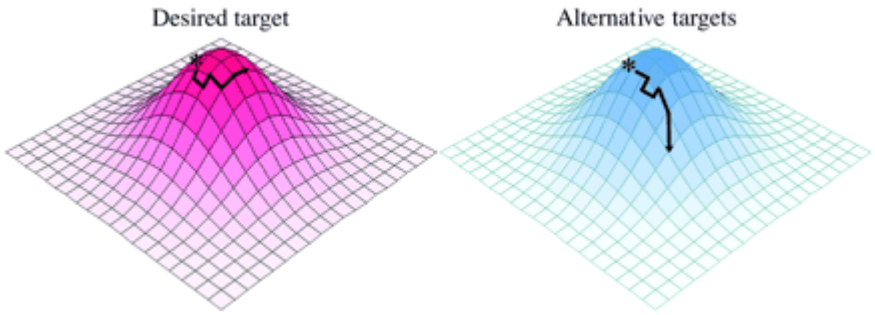


Figure 1.2: "(A) X-ray structure of CaM in complex with smMLCK (PDB code 1CDM) generated with MOLMOL. Ca^{2+} atoms are shown as yellow spheres. The CaM surface and boundary residues selected for optimization are shown in red. The core residues in the CaM-binding interface, shown in cyan, and the smMLCK residues, shown in light yellow, were allowed to change conformation during the optimization procedure.(B) Fitness of CaM interacting with the desired target, smMLCK (red), and with alternative targets (blue). The CaM_{wt} sequence lies near the respective maxima for both the desired target and the alternative targets as indicated by *. Arrows show the change in fitness due to mutations in the CaM sequence predicted by the optimization." Figure and caption reproduced from Shifman, JM Mayo, SL Proc. Natl. Acad. Sci (2003), 100, 13274.

1.4 Directed evolution

Although rational re-design of protein substrate specificity is a powerful tool for enzymes with well characterized structures and substrates, the method of choice to generate novel enzymes with altered functions remains directed evolution. The advantage of using directed evolution is that it is a combinatorial approach to screen for the relevant phenotypic function. The use of directed evolution to evolve modified proteins necessitates the use of either a genetic selection to select for survival or a screen/assay to scan libraries. Basically, directed evolution (Figure 1.3) involves three steps; first the candidate enzyme(s) to be evolved is identified; second, a combinatorial library of enzyme mutant genes is generated and third a screen/selection of enzyme activity is employed to pick out winners. Additional rounds of mutagenesis/screening can be performed until the enzyme variant has the desired new functional activity. A recent review on directed evolution summarizes most of the current methodologies to generate diversity at the genetic level (20) including but not limited to, error-prone PCR (ePCR), DNA shuffling, incremental truncation for the creation of hybrid enzymes (ITCHY) and staggered extension protocol (StEP).

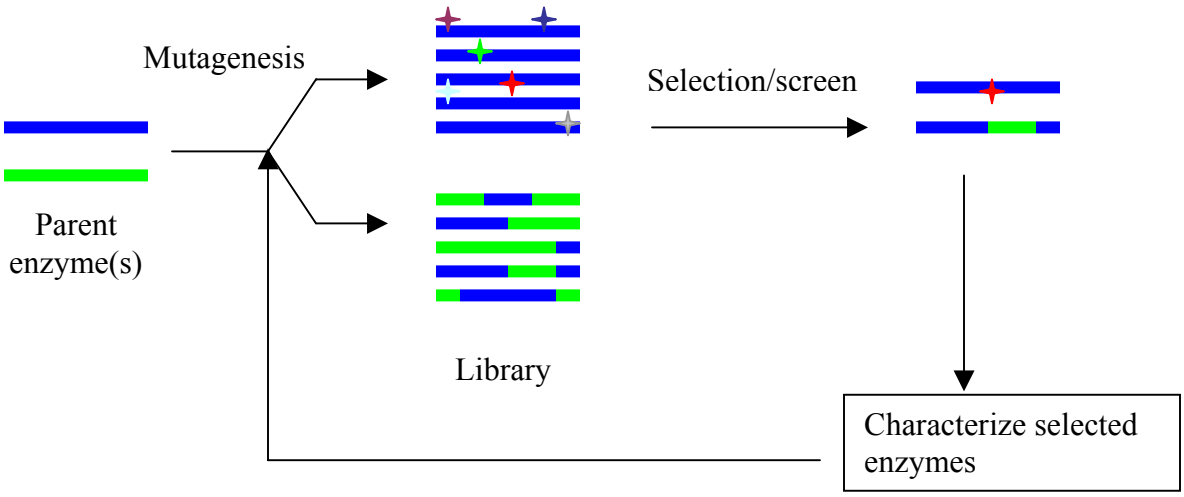


Figure 1.3: Directed evolution scheme. Two different mutagenesis strategies are shown; error-prone PCR indicated by colored stars on a single gene framework; and recombination obtained by crossing-over one or more genes.

1.4.1 Restriction endonucleases/recombinases.

Restriction endonucleases are the backbone of recombinant DNA technology and genome engineering (21). Restriction enzymes represent a particularly stringent model system for altering enzyme substrate specificity since any engineered enzyme must recognize the requisite target DNA sequence in the context of a wide variety of other sequences including the DNA sequence recognized by the starting parent(s). Plus, there are detailed structures of restriction endonucleases complexed with cognate as well as non-cognate substrates (22). Arnould *et al.* use a semi-rational approach to alter the DNA specificity of I-CreI, a homing endonuclease of the LAGLIDGAG family (23). First, they

constructed a lacZ reporter plasmid in which the lacZ gene is interrupted with an insert containing the site of interest flanked by two direct repeats. Next, a strain harboring the expression vector encoding the endonuclease library was mated with the strain containing the interrupted lacZ reporter (Figure 1.4). Upon mating, active endonucleases restore functional lacZ by performing double stranded breaks at the sites of interest. Four variants that had altered specificity were extensively characterized *in vitro* and displayed half-maximal reaction kinetics comparable to the wild-type (WT) parent. Additionally, since the LAGLIDGAG family of endonucleases normally function as homodimers, variants that cleaved new DNA targets can be assembled by creating heterodimers. Although cleavage at sequences recognized by either homodimers still occurred, the authors showed that hybrid DNA sequences could now be recognized.

Working with the same LAGLIDADG family of endonucleases, Chevalier ***et al.*** used a combination of computational redesign and an *in vivo* folding screen to generate heterodimeric chimeras of I-DmoI and I-CreI (24). The best enzyme, E-DreI16, has eight computationally designed point mutants at the domain interface and recognizes a hybrid DNA sequence not recognized by either parent, but shows no cross-reactivity to the wild-type preferred DNA sequences.

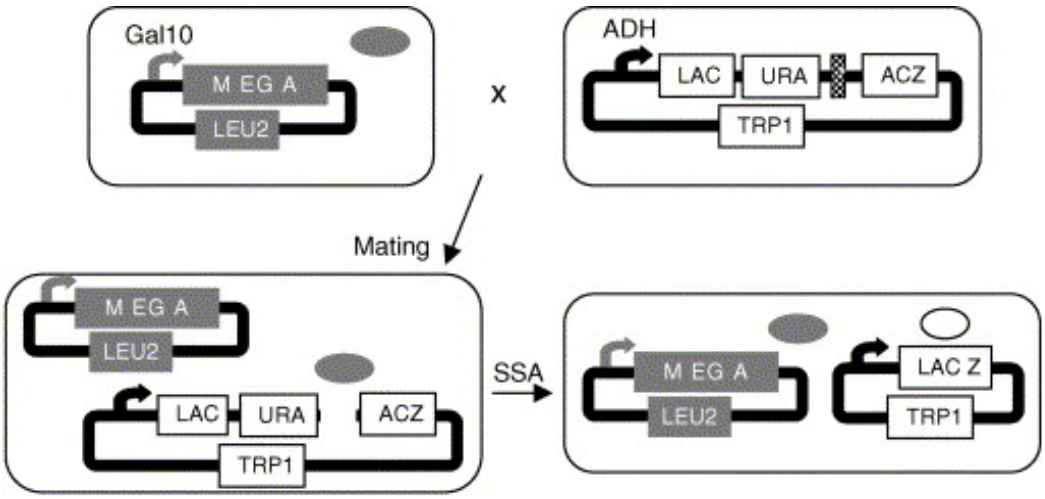


Figure 1.4: "The yeast mating assay for the selection of modified I-CreI variants." Figure and caption reprinted from Arnould *et al.* (2006) *J. Mol. Biol*, **355**, 433. Copyright © 2006, with permission from Elsevier.

Samuleson *et al.* use a dual genetic selection protocol to isolate NotI mutants that recognize the alternate DNA sequence 5' GC**T**GCCGC 3' (25). In step 1, an SOS induction assay was employed by using a DNA-damage indicator strain ER1992 harboring a pACYC-eagIM plasmid to pre-modify and protect all 5' CGGCCG 3' sites within the genome. Endonuclease cleavage at all sites other than 5' NCGGCCGN 3' would induce the SOS response leading to a lacZ readout. By screening a randomly mutated *notIR* gene library, they isolated the E156K mutant that cleaved at 5' GC**T**GCCGC 3'. In order to further improve the activity of the variant, a second selection was performed wherein highly active mutants would eliminate a vector harboring a toxic

protein tox-yidC. The best mutant, M91V/E156K, was able to cleave seven alternate sites with nearly the same efficiency as the cognate NotI site, indicating a relaxation in the specificity of the enzyme.

BsoBI is a thermophilic restriction endonuclease that recognizes the DNA sequence 5' CPyCGPuG 3'and cleaves after the first cytosine. According to the structure of the BsoBI-DNA co-crystal, the Asp246 of BsoBI makes a water-mediated hydrogen bond to N6 of the degenerate base adenine. Zhu *et al.* used saturation mutagenesis at Asp246 to look for BsoBI variants with altered DNA cleavage specificity (26). The Asp246Ala mutant had a 70-fold greater affinity for the 5' CCCGGG 3' DNA sequence. Using an SOS induction assay, they screened for revertants that cleaved the wild-type (WT) preferred sequence 5' CTCGAG 3' and found that most of the mutations that restored WT-like activity were located outside the DNA-protein interface.

Two different groups have independently designed a similar selection scheme for altering the specificity of the restriction endonuclease I-SceI by modulating survival of cells harboring the "control of cell death B" (CcdB) protein (27, 28). By incorporating a simultaneous negative selection to weed out non-specific mutants, Doyon *et al.* report the isolation of 2-4, an I-SceI variant that showed comparable *in vivo* activity to the WT I-SceI enzyme (28).

FokI, a type IIS restriction enzyme**,** is a broad non-specific DNA cleaving enzyme. The C2H2 zinc finger is a common DNA binding motif. Engineered zinc fingers (29) with novel specificities can be coupled to the cleavage domain of FokI to generate zinc-finger

nucleases (30). Urnov *et al.* have employed a GFP-based flow-cytometric gene correction assay to engineer zinc-finger nucleases designed against the X-linked severe combined immune deficiency (SCID) in the IL2Rg gene (31).

Site-specific recombination of DNA (32) is catalyzed by recombinases that perform three essential functions; recognition of the target sites, cleavage, and rejoining. The Cre recombinase from bacteriophage P1 recognizes a 34bp double stranded DNA sequence called loxP. Two independent reports demonstrate the engineering of the Cre recombinase to recognize alternate DNA sequences. Buchholz *et al.* used a recombination based PCR strategy to identify mutant Cre recombinases that cleaved an alternate DNA seuence termed loxH (33). In order to eliminate non-specific mutants that recognized both loxH and loxP, they performed an additional selection to eliminate cross-reactivity. The most specific recombinase mutant, Fre22, showed selectivity for loxH over the WT loxP sequence, as determined by both *in vivo* and *in vitro* experiments. Santoro *et al.* used a fluorescence-activated cell sorting (FACS) based approach for identifying Cre mutants that selectively recognize unnatural recombination sites (34). They used a two plasmid system, one containing a Cre recombinase contact residues library and the other containing two fluorescent proteins GFPuv and EYFP. Cells containing an active recombinase would rearrange the reporter plasmid leading to GFP and EYFP fluorescence that are monitored in independent channels of a flow-cytometer. By performing additional rounds of negative selection where recombination is based on

the natural loxP sites, they isolated Cre(+-)#4 that showed selectivites and *in vitro* recombination frequencies comparable to that of the WT protein.

One of the paradigms of directed evolution is, " you get what you screen for". Wherever enzyme specificity is a strict requirement and cross-reactivity does not lead to a desirable enzyme, as illustrated with restriction enzyme engineering above, the incorporation of simultaneous negative selections facilitates the isolation of specific enzyme variants.

1.4.2 Engineering enzyme substrate specificity for organic synthesis

The enantioelective synthesis of organic compounds/pharmaceuticals is a billion dollar industry constantly requiring the use of steroselective and efficient catalysts (2, 35). Enzymes are ideally suited for this purpose not only because they are extremely efficient catalysts but also because most of them are steroselective. Unfortunately, the use of enzymes is limited by their narrow substrate selectivity and lack of activity in organic solvents. A substantial amount of research has then been focused on altering the specificities of natural enzymes to tune them to commercial substrates and towards increasing their activities in organic solvents. Gupta *et al.* have recently reviewed the use of enzymes in organic solvents and their engineering in low water content media (3). In the following section, a few notable examples of enzymes engineered for organic synthesis are presented.

Cytochromes P_{450} constitute a superfamily of enzymes capable of catalyzing a diverse array of organic transformations. One of the most challenging problems in organic

chemistry is the activation of the relatively strong C-H bond. Dioxygen-supported alkane hydroxylation can be achieved though the use engineered cytochrome P_{450}s (36, 37). Although the preferred substrates of cytochrome P_{450} BM3 are medium chain fatty acids (C12-C18), Peters *et al.* have engineered P_{450} BM-3 variants that can catalyze the oxidation of very small alkanes like methane and ethane to the corresponding alcohols (38). Interestingly, two of the isolated variants were found to be highly regioselective and eneantioselective towards small chain alkanes (C7/C8).

Urlacher *et al.* have summarized the applications of microbial P_{450}s in biotechnology (39, 40). Other efforts towards bio-hydroxylation though substrate engineering are highlighted in a recent review by Raadt & Griengl (41).

Aldolases are enzymes that catalyze enantioselective C-C bond forming reactions. Despite their widespread potential, narrow substrate specificity limits their use in industrial scale processes. Sialic acid aldolase catalyzes the condensation of *N*-acetyl-D-manosamine and pyruvate to yield *N*-acetyl-D-neuraminic acid (sialic acid). Hsu *et al.* engineered the Sialic acid aldolase to invert the stereochemistry of the substrate sugar (Figure 1.5) to accept L-3-deoxy-manno2-octulosonic acid (L-KDO) (42). By coupling the aldol cleavage reaction to the reduction of pyruvate to lactate and the concomitant oxidation of NADH to NAD+ by lactic dehydrogenase, they assayed for aldolase variants by monitoring NADH fluorescence. After five rounds of screening, they isolated aldolase mutant N5B2 that showed a completely different sugar specificity profile, with L-KDO being the preferred substrate.

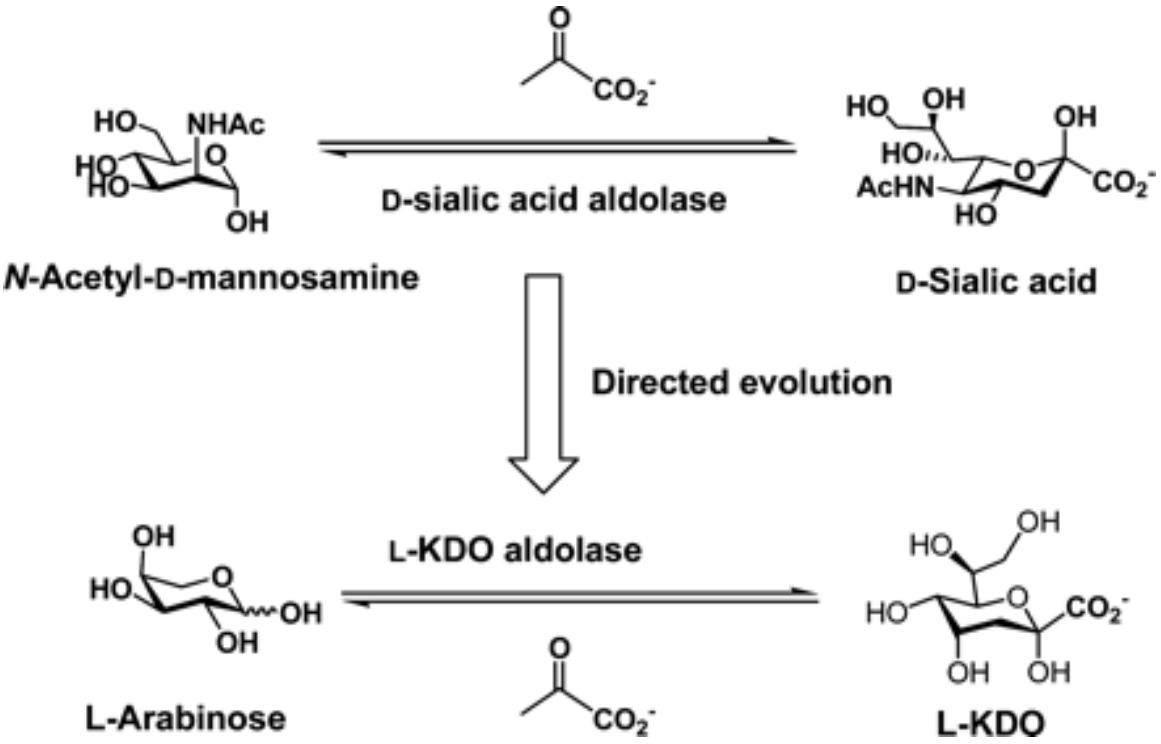


Figure 1.5: "Directed evolution of D sialic acid aldolase to L-KDO aldolase." Figure and caption reproduced from Hsu *et al.* Proc. Natl. Acad. Sci. (2005), 102, 9122. Copyright © 2005 by the National Academy of Sciences.

An orthogonal approach to engineer aldolases is to alter the stereochemical course of the bond-forming step, enabling the formation of a completely different diastereomer from the same starting materials. Fructose-1,6-biphosphate aldolase (FBPA) catalyzes the formation of fructose-1,6-biphosphate starting with dihydroxyacetone phosphate (DHAP) and glyceraldehydes-3-phosphate (G3P). Williams *et al.* used a DNA shuffling (43)

based approach to isolate mutants of the D-tagatose-1,6-bisphospate (TBPA) that could catalyze the formation of the FBP (44). The best variant, 3-23B5, had only four mutations relative to TBPA, but still showed a ~70 fold improvement in activity towards FBP.

The different classes of aldolases and their substrate preferences have been summarized in a recent review by Samland & Sprenger (45). Franke *et al.* have highlighted the recent directed evolution efforts towards aldolase variants (46).

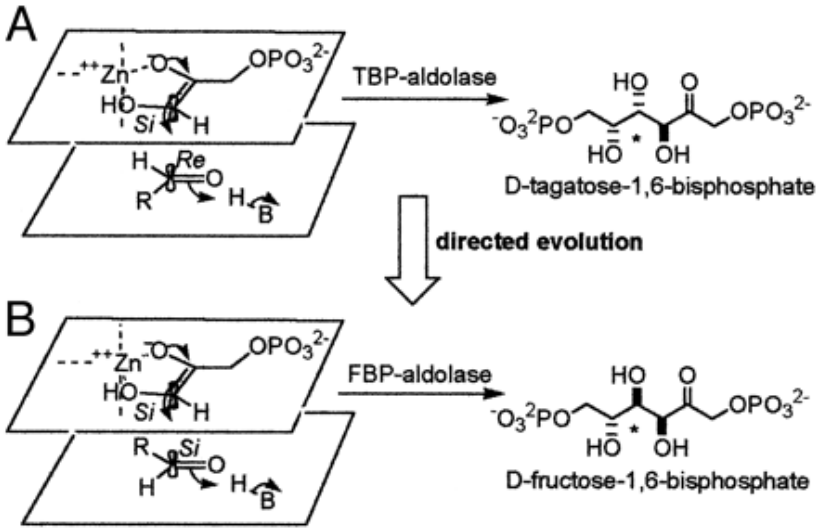


Figure 1.6: Stereochemistry of the reaction catalyzed by aldolases. (*A*) The mechanism of TBP aldolase. The DHAP ene-diolate is formed after abstraction of the 1-*proS* proton from DHAP and polarization by the catalytic zinc cation. Attack of the activated DHAP C1 from its *Si* face onto the G3P C1 *Re* face generates the 3*S*, 4*S* product tagatose 1,6-bisphosphate, and proton donation by H-B (Asp-82) converts the C4 carbonyl to a hydroxyl group, completing TBP synthesis. (*B*) The mechanism of FBP aldolase. The DHAP ene-diolate is formed after abstraction of the 1-*proS* proton from DHAP by Glu-182 and polarization by the catalytic zinc. Attack of the activated DHAP C1 from its *Si* face onto the G3P C1 *Si* face and proton donation by H-B (Asp-109) convert the C4 carbonyl to a hydroxyl group, completing the synthesis of the 3*S*,4*R* product fructose 1,6-bisphosphate. FBP and TBP are epimeric at C4, and this position is marked with an asterisk. R = $CH(OH)CH_2OPO_3^{2-}$. Figure and caption reproduced from Williams *et al.* Proc. Natl. Acad. Sci. USA (2003) 100, 3143. Copyright © 2003 by the National Academy of Sciences.

Asymmetric catalysis is one of the foremost challenges in organic synthesis. Using directed evolution to modify the enantioselectivity of enzymes, a variety of transformations have now been performed to yield enantiopure products (47). Natural product glycosides such as calicheamicin, doxorubicin, vancomycin and nystatin have therapeutic applications. The chemoenzymatic synthesis of such glycoconjugates often depends on the availability of the corresponding sugar-1-phosphates. Starting with the *E. coli* galactokinase GalK that recognizes D-galactose, Hoffmeister *et al.* used a high-throughput multi-sugar colorimetric screen to isolate a GalK mutant Y371H that displayed a substantial degree of kinase activity towards L-sugars like L-glucose and L-altrose (48). Surprisingly, the Y371H mutant could still phosphorylate D-galactose efficiently, making it an anomeric D/L sugar kinase.

Reetz *et al.* used a medium-throughput GC based screening system to derive enatioselective variants of cyclohexanone monooxygenase (CHMO) as catalysts in the Baeyer-Villiger oxidation of 4-hydroxycyclohexanone (49). Although the WT enzyme favored the formation of R-lactone with an enantiomeric excess (ee) of just 9%, the best variant, 1-H7-F4, favored the formation of the R lactone with an ee of 54%. Another variant, 1-K2-F5, identified in the same screen favored the formation of the S lactone with an ee of 79%. CHMO enzyme variants were thus evolved that were specific for the formation of either enantiomer. The same group also reported the evolution of CHMO variants that catalyze the air oxidation of thioesters to yield either the R or the S enantiomer with >95% ee (50).

The application of high-throughput assays for the directed evolution of lipases and esterases has recently been reviewed by Schmidt & Bornscheuer (51).

A completely different methodology for the creation of enzymes that catalyze organic transformations is the engineering of artificial metalloenzymes. The methodology involves grafting of coenzyme analogs or catalytic metal co-factors onto the active site of protein scaffolds either using site-specific, covalent anchoring (52), or supramolecular anchoring (53). A recent review by Thomas & Ward succinctly summarizes the field (54).

1.4.3 Acylases/ Proteases

Proteases and acylases are enzymes that catalyze the hydrolytic cleavage of amide bonds. Proteases/peptidases cleave the amide bonds in proteins/peptides and are involved in the degradation and activation of proteins. Acylases on the other hand hydrolyze amide bonds other than those found in peptides.

Semi-synthetic cephalosporins are an important class of commercial antibiotics. 7-aminocephalosporanic acid (7-ACA) and 7-aminodesacetoxycephalosporanic acid (7-ADCA) are important intermediates in chemoenzymatic synthesis of cephalosporins. Sio *et al.* performed site-saturation mutagenesis at residue Phe375 to investigate the ability of the *Pseudomonas* SY-77 gluataryl acylase to perform enzymatic deacylation of adipyl-7-ADCA to yield 7-ADCA (55). Although the k_{cat} for the SY-77Phe375Cys variant with 7-ADCA was the same as wild-type, the K_M decreased by about 6-fold to yield an overall

6-fold more efficient enzyme. Not surprisingly, almost all of the variants had a preference for the WT substrate gluataryl-7-ACA. The same group had earlier published a paper on the directed evolution of β-subunit of the cephalosporin acylase to alter the substrate specificity towards 7-ADCA (56). Using 7-adipyl leucine as the sole leucine source on minimal plates, they isolated variants that were 8-fold better than the starting parent in catalyzing the deacylation reaction. Sio & Quax have also reviewed the industrial applications of β-lactam acylases (57).

In an interesting application of the directed evolution of substrate specificity, Weinreich *et al.* showed that Darwinian evolution of cefotaxime resistance in the TEM family of β-lactamases could follow very few mutational pathways (7). In an earlier study, Orencia *et al.* had applied directed evolution to predict the emergence of antibiotic resistance (6). Working with a hypermutator strain of *E. coli* they evolved cefotaxime resistance based on a survival selection and showed that the *in vitro* engineered variant Glu104Lys/Met182Thr/Gly238Ser (Figure 1.7) was identical to the clinical isolate TEM-52.

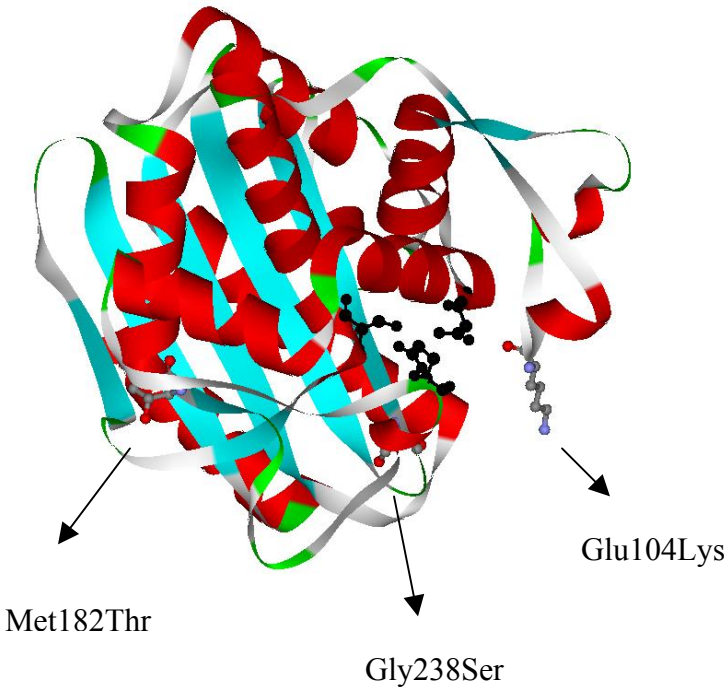


Figure 1.7: The crystal structure of the TEM-52 clinical isolate (PDB ID: 1HTZ). The active site residues are shown in black.

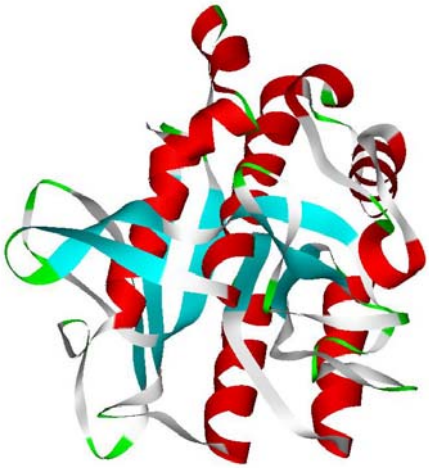
Proteases are ubiquitous proteins that play a key role in a wide range of physiological processes like protein degradation, signaling, apoptosis, regulation and activation. Consequently, a plethora of proteases exist with varying specificities and activities. One of the most well known examples of protease substrate engineering is the conversion of trypsin into chymotrypsin. Although both proteins are serine proteases and share considerable structural homology, chymotrypsin cleaves after large hydrophobic residues

while tryspin cleaves after basic residues. Converting trypsin into a chymotrypsin like protease, however, required the modification of entire surface loops 185-188 and 221-225 (58).

Alkaline proteases have long generated industrial interest because of their application as detergent additives. Gupta *et al.* have summarized the optimization and applications of bacterial alkaline proteases for industrial applications (59) and Maurer has reviewed detergent proteases (60).

Subtilisins are a family of alkaline serine proteases secreted by a wide variety of Bacillus species. Subtilisins (Figure 1.8) represent one of the most well studied, characterized and engineered enzymes owing to their large-scale industrial applications as detergent additives and skin and leather processing enzymes. The enzyme has no disulfide bonds and consists of a single polypeptide chain of about 275 residues. Subtilisin is synthesized as the apo-enzyme (a characteristic feature of proteases), secreted across the cytoplasmic membrane and matured into the final active form by cleavage of the pro-sequence by autocatalysis. The pro-sequence is believed to mediate the folding of the final mature form of subtilisin (61). Upon completion of folding the pro-sequence is then auto-catalytically degraded (62). Bryan has recently reviewed the engineering of subtilisin (63).

A



B

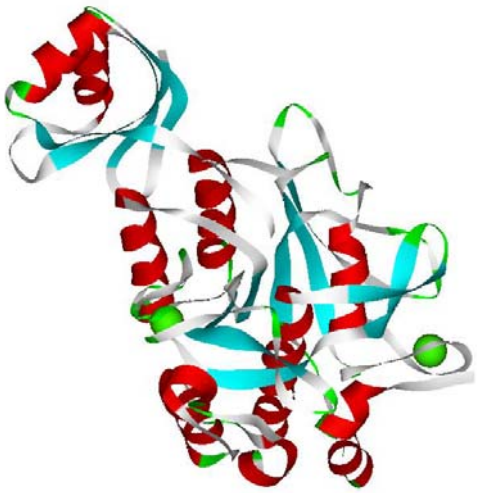


Figure 1.8: The crystal structures of (A) subtilisin BPN (PDB ID: 1GNS) and (B) autoprocessed Ser221Cys-subtilisin E-propeptide complex (PDB ID: 1SCJ)

The human immunodeficiency virus type I protease (HIV PR) catalyzes the cleavage of nine divergent amino acid sequences within the HIV-1 Gag-Pol, thereby enabling the activation of proteins essential for viral replication. The conformation and shape of the polypeptide being cleaved as opposed to consensus amino acid sequences dictates recognition. It is essentially a non-specific protease and is cytotoxic when expressed in either mammalian or *E. coli* cells. By using a high-throughput β-gal inactivation assay O'Loughlin *et al.* selected for HIV PR mutants that were non-toxic to the *E. coli* cells in which they were expressed (64). One variant, P9S/I150L, exhibited improved specificity for TNF-α/βGAL, the selection substrate. In addition, the double mutant showed reduced cytotoxicity when expressed in *E. coli* cells. Western blotting and *in vitro* analysis using a synthetic HIV-1 peptide confirmed the specificity inversion.

The *E. coli* outer membrane endopeptidase OmpT is a member of the omptin family of β-barrel proteins. The enzyme has a strong preference for cleavage at dibasic residues (65). Olsen *et al.* from our group used a fluorescence-activated cell sorting (FACS) based screen in order to isolate OmpT variants (Figure 1.9) (66). The selection peptide was a FRET peptide with a positively charged tail that was used to anchor the fluorescent moiety upon proteolysis onto the negatively charged *E. coli* outer surface. By screening an error-prone library OmpT variants were isolated that cleaved an Arg-Val sequence. The best variant, C5, showed a 60-fold improvement in catalytic efficiency for cleavage of an Arg-Val sequence compared to OmpT.

The two papers highlighted above are rare examples of protease screening *in vivo*. Traditionally, high-throughput protease screening *in vivo* has been challenging due to two main obstacles, cytotoxicity, and auto-proteolysis. The isolation of highly active soluble protease variants *in vivo* seems unlikely, mainly to due to the degradation of essential proteins. Similarly, highly active proteases have a tendency to cleave themselves leading to loss of function. Interestingly, many proteases in nature, including chymotrypsin and trypsin, use this auto-proteolytic function to achieve the opposite effect. These proteases are synthesized as their zymogens and auto-proteolysis leads to activation of protease function.

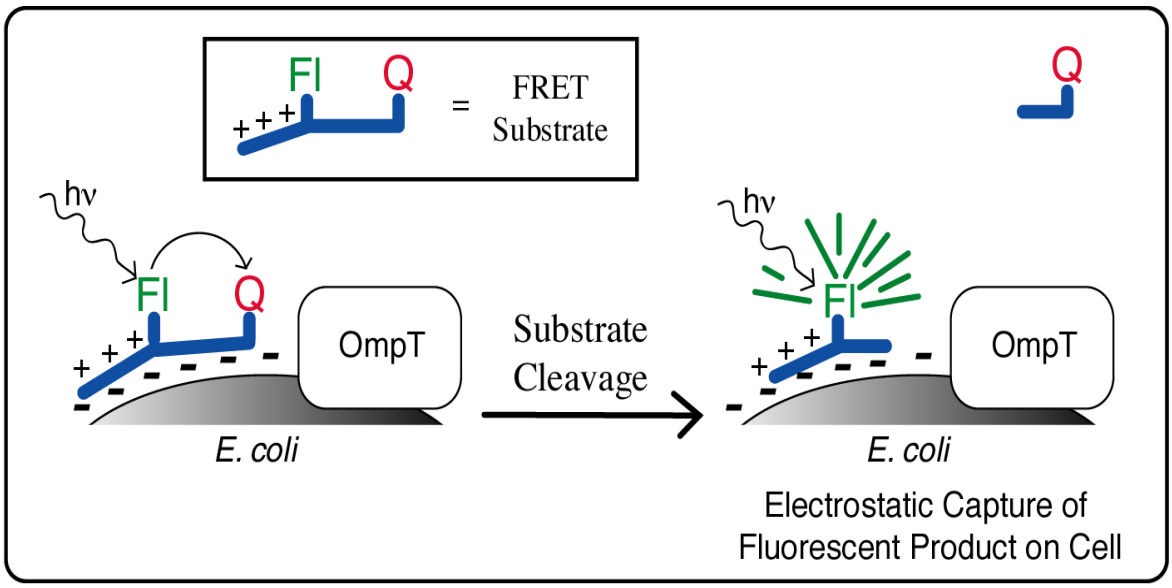


Figure 1.9: Flow-cytometric assay used for the screening of OmpT libraries expressed on the surface of *E. coli*.

D-amino-peptidase (DAP) and the R61 D,D-carboxypeptidase are members of the large family of penicillin-recognizing, active site serine enzymes. D-aminopeptidase catalyzes the hydrolysis of *N*-terminal D-Ala residues while the R61 D,D-carboxypeptidase (R61) catalyzes the hydrolysis of D-Ala from peptide fragments containing two *C*-terminal D-Ala residues. The DAP enzyme is comprised of three distinct domains A, B &C. The backbone of the DAP-A domain harbors the active site and is structurally homologous to the R61 enzyme. By using a combination of homology structure guided site-directed mutagenesis and γ-loop deletion, Delmarcelle *et al.* engineered a mutant DAP-A domain (DAP-475G-487-N275R) that when complemented with B&C domains showed carboxypeptidase activity *in vitro* (67).

Domain recombination allows for assembly of structurally independent functional domains. In proteins where the catalytic site and association/folding are well separated, domain recombination can generate new functions. Factor Xa protease and trypsin belong to the S1 serine protease family whose two homologous β-barrel subdomains assemble to form the binding sites and catalytic machinery. Hopfner *et al.* engineered a factor Xa (fXa)/trypsin chimera by combining the *N*-terminal β-barrel of fXa with the *C*-terminal β-barrel of trypsin (68). The hybrid protein, rfXYa, showed a substrate specificity profile that was different from either starting parent.

1.4.4 Transferases

Transferases are enzymes that catalyze the transfer of functional groups from a donor molecule to an acceptor molecule. The important classes of enzymes from a directed evolution standpoint in the transferase family are glycosylases, transaminases (aminotransferases) and glutathione transferases.

Glutathione-S-transferases (GSTs) are essential components of the cellular detoxification machinery and eliminate reactive electrophilic compounds by conjugating them to the Glutathione (GSH) tripeptide. The theta class GSTs posses a *C*-terminal alpha helical extension that contains both the electrophile and GSH binding sites in addition to the highly conserved GST fold. Griswold *et al.* used a combination of homology-dependent and homology-independent methodologies for generating chimeric GSTs starting with the rat GSTT2-2 (rGSTT2-2) and the human GSTt-1-1 (hGSTT1-1) (69). They screened the chimeric GST library using a high-throughput flow-cytometric assay for isolating variants that had high levels of rGSTT2-2 like activity (Figure 1.10). One of the isolated clones, SCR23, substituted two H-site helices from rGSTT2-2 into the hGSTT1-1 framework and had a catalytic efficiency that was 3.5 times better than the rGSTT2-2 and 300 times better than the hGSTT1-1. Additionally, SCR23 conjugated GSH to ethacrynic acid, a reaction disfavored by both starting parental enzymes. The ability to shuffle enzymes with low sequence homology demonstrated in the preceeding example should help accelerate the discovery of humanized enzymes with low immunogenicity. The same group has also reported the directed evolution of rGSTT2-2 like activity starting with

hGSTT1-1 by using a combination of error-prone PCR and DNA shuffling (70). The best variant, HEPShEP-aa7, showed a 20,000-fold increase in activity compared to the starting parent, hGSTT1-1, for the conjugation of CMAC to GSH. Interestingly, construction of a three-residue saturation library (the residues were identified based on other screening experiments) and subsequent flow-cytometric screening led to the isolation of S3-7, an enzyme with ~4000-fold increase in reactivity with CMAC.

Emren *et al.* constructed chimeric GST libraries by shuffling the cDNA encoding the human GSTM1-1 (M1) and GSTM2-2 (M2) genes (71). A set of 384 random clones from the library and the two parental enzymes were expressed individually in *E. coli*. The multi-dimensional catalytic activities were analyzed using a set of eight different substrates. Using principal component analysis (PCA) to analyze this multidimensional space they identified two enzyme variants, 342 &383, that had substrate reactivity profiles different from M1 and M2.

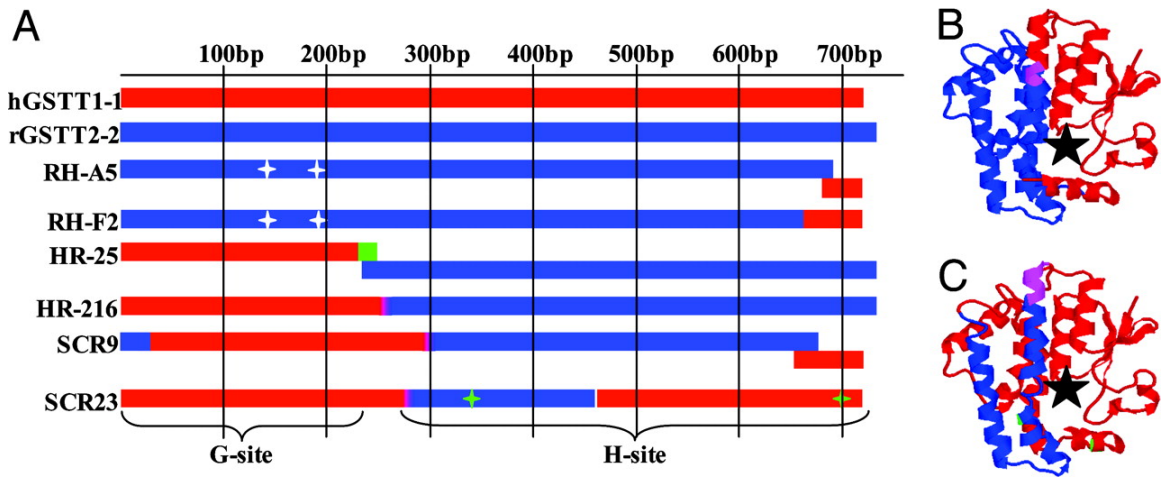


Figure 1.10: "Gene and protein structures of GSTT enzymes. (*A*) Schematic representation of selected chimeric genes. Sequences inherited from hGSTT1-1 are shown as red bars, and segments from rGSTT2-2 are shown as blue bars. Insertions not derived from either parent are represented as green bars. Positioning of the progeny segments corresponds to their origin in the parental genes (depicted at top). Point mutations are represented as white (silent) or green (encoding for amino acid substitution) stars. Sequences encoding the G and H sites are noted. (*B*) Mapping of SCR9 amino acid sequence onto the crystal structure of the hGSTT2 monomer. Human-derived sequence is in red, rat sequence is in blue, and identical amino acids at fusion points are in magenta. The location of the active site is marked with a black star. (*C*) Mapping of SCR23 amino acid sequence onto the hGSTT2 structure. Point mutations are shown in green." Figure and caption reproduced from Griswold *et al.* Proc. Natl. Acad. Sci. USA (2005), 102, 10082. Copyright (c) 2005 by the National Academy of Sciences.

Aminotransferases are essential metabolic enzymes that catalyze the transfer of an amino group between amino acids and 2-oxo acids. The aspartate 2-oxoglutarate aminotransferase (AspAT) from *E. coli* displays a marked preference for acidic substrates and shows almost no activity with β-branched amino acids like valine. Yano *et al.* used a selection system based on growth of auxotrophic *E. coli* complemented with AspAT libraries to evolve AspATs that catalyze the transamination of 2-oxovaline (72). They first knocked out the branched-chain amino acid aminotransferase gene, *ilvE* from *E. coli*. An *ilvE* deficient strain cannot grow on minimal plates without being supplemented with valine, leucine and isoleucine. Next, they transformed an AspAT shuffled library into the knockout strain so that AspAT mutants that showed high efficiency towards β-branched amino acids could complement the *ilvE* knockout and lead to survival on minimal plates. Using this selection scheme, they isolated a mutant AspAT enzyme, AV5A-7, that catalyzed the transamination of 2-oxovaline 10^5 times better than the starting parent enzyme. To date, this represents one of largest increases in activity towards a non-natural substrate for an engineered enzyme.

In a series of papers, the Kirsch group has reported the directed evolution of the *E. coli* aspartate aminotransferase (eeAATase) towards converting it into the homologous tyrosine aminotransferase (eeTATase). Using homology modeling, they first constructed HEX by introducing substitutions rationally based on the active site of eeTATase (73). Although the activity of HEX towards phenylalanine increased by more than three orders of magnitude, the enzyme still preferred aspartate. Starting with HEX as a template, they

made a library of HEX mutants using DNA shuffling and selected for variants that complemented tyrosine auxotrophy (74). This led to the construction of the HEX+ Ala293Asp variant that preferred tyrosine to asparate, by a factor of six. A third eeTATase variant, SRHEPT, constructed by combination of selection and rational mutagenesis, showed an activity and specificity profile similar to HEX (75). A particularly interesting aspect of the research is the comparison between the *in vitro* evolved enzyme and its natural counterparts. By incorporating the Ala293Asp mutation into SRHEPT, Chow *et al.* further increased the Phe/Asp specificity of the enzyme variant (76).

Glycosyltransferases are enzymes that catalyze the transfer of activated monosaccharides onto a wide variety of acceptor substrates. These enzymes catalyze the synthesis of all carbohydrates. Recently, two different approaches have been reported for the directed evolution of glycosyltransferases. Aharoni *et al.* have developed a flow-cytometric assay for the engineering of sialyl transferases (77). They used a two-plasmid system, one carrying the CMP-Neu5Ac synthetase gene to activate Neu5Ac (sialic acid) and the other plasmid encoding the CstII sialyl transferase library. Cells expressing active CstII variants that could conjugate sialic acid to BODIPY lactose (fluorescent acceptor sugar) would thus be fluorescent, allowing isolation via a fluorescence activated cell sorter (FACS). Using this strategy, they isolated CstIIPhe91Tyr, a mutant that showed a 153-fold improvement in catalytic efficiency relative to the WT enzyme in utilizing BODIPY-lactose as the acceptor sugar. In addition, the CstIIPhe91Tyr variant could also utilize

BODIPY-3SH-lactose as an acceptor, a sugar not recognized by the WT enzyme. Love *et al.* describe a phage display (For a review of phage display for the directed evolution of enzymes please see (78)) based assay for the evolution of glycosyltransferases (79). They used a phagemid to express the *E. coli* glycosyltransferase (MurG) as a fusion to pIII. and The phage-bound MurG to catalyze the transfer of ^{14}C-labeled UDP-GlcNac onto a biotinylated lipid which was captured on a streptavidin coated membrane. Scintillation counting was used to quantitate the formation of the lipid conjugate.

Glycosidases have the opposite effect of glycosyltransferases; they catalyze the hydrolysis of glycoside bonds. Feng *et al.* describe the engineering of β-glycosidase of *Thermus thermophilus* towards increasing its ability to catalyze the transglycosylation reaction (80). They constructed an error-prone library of the β-glycosidase (β-Gly) enzyme and used a negative screen to identify mutants that displayed low levels of hydrolytic activity. Working with this smaller subset of variants, they selected for enzymes that showed a high transglycosylation activity to yield enzymes with favorable transglcysoylation/hydrolysis ratios. The best enzyme, Phe140Ser, could effect regiospecific transglycosylation while simultaneously suppressing glycolysis. Given the difficulties in oligosaccharide synthesis and the challenges associated with glycosyltransferase substrate engineering, directed evolution of soluble glycosidases appears a viable alternative towards assembling these sugars.

There are numerous examples of the engineering of substrate specificity of glycosidases towards hydrolysis of alternate sugars. Highlighted below, is one such example. A recent

review by Hancock *et al.* summarizes the engineering of glycosidases and glycosyltransferases (81).

Beta-glucuronidase catalyzes the hydrolytic conversion of β-D-glucuronoside to D-glucuronate. Since the *E. coli* beta-galactosidase (lacZ) and the beta-glucoronidase (gusA) probably diverged from a common ancestor, Mastsumura & Ellington engineered a gusA enzyme variant that could catalyze the hydrolysis of β-galactoside (82). Using the standard X-gal plate assay, they isolated a quadruple mutant Thr509Ser/ Ser557Pro/ Asn566Ser/ Lys568Gln that cleaved a β-galactosidase substrate 500 times more efficiently than the wild-type enzyme. Additionally, the gusA mutant preferred the pNP-galactoside substrate to the pNP-glucoronide substrate.

1.4.5 Bioremediation enzymes

One of our key weapons in the fight against organic/toxic pollution is bioremediation using either whole microbial organisms or their enzymes. As a byproduct of environmental contamination, bacteria and other microorganisms have evolved to use these pollutants/contaminants as a source of nutrients in a remarkable illustration of evolution/adaptation. The ability to engineer enzyme substrate specificity of these bioremediation enzymes enables the detoxification of a wider array of contaminants.

Organophosphate pesticides like parathion and chlorpyrifos and chemical warfare nerve agents like soman and cycolosarin are extremely toxic by virtue of their irreversible inhibition of acetylcholinesterase. The human serum paraoxonase (PON1) is primarily a

lactonase but is also capable of hydrolyzing organophospates like paraxon. PON1s are enzymes with broad substrate specificity and are expressed poorly in *E. coli*. Aharaoni *et al.* used DNA family shuffling based solubility engineering (83) to increase the recombinant expression of the PON1s in *E. coli* (84). Three rounds of shuffling followed by screening lead to the isolation of variant G3C9 that could be expressed at 12 mg/l in *E. coli*. Since the G3C9 expression variant, like its parent PON1, demonstrated greater catalytic efficiency towards hydrolysis of phenylacetate than organophosphates, the authors constructed a random mutagenic library of G3C9 variants to select for enzymes that preferred a coumaphos like the fluorogenic substrate DEPCyC. Screening resulted in the isolation of two new enzymes, G3C9.10 and G3C9.49 that showed a preference for DEPCyC hydrolysis compared to phenylacetate hydrolysis. They also used family shuffling to increase the rate of hydrolysis of paraoxon 240-fold in the homologous enzyme PON3.

1,2,3-trichloropropane (TCP), a member of the family of chlorinated hydrocarbons, is a toxic water pollutant. Chlorinated hydrocarbons are widely used a solvents in organic synthesis. Haloalkane dehalogenases are α/β-hydrolases that catalyze the conversion of haloalkanes to the corresponding alcohols. The haloalkane dehalogenase (DhaA) from *Rhodococcus* is an enzyme with a broad specificity. Since the byproduct of the dehalogenase reaction is the corresponding hydrohalide, Bosma *et al.* used a plate-screening assay based on the pH indicator eosin-methylene blue (EMB) (85). A 10,000 member randomly shuffled library was screened for the development of red color after

exposure to TCP vapor. After two rounds of screening they isolated a DhaA variant M2 that showed ~8-fold improvement in hydrolyzing TCP. In order to test the feasibility of the evolved mutant in eliminating TCP, the M2 variant was introduced into the 2,3-dichloro-1-propanol utilizing bacterium *Agrobacterium radiobacter*. Indeed, this engineered strain could now metabolize TCP and the pH of the culture increased as result of the production of hydrochloric acid. Janssen has recently summarized the efforts towards the evolution of haloalkane dehalogenases as bioremediation enzymes (86).

Nitroaromatics are a class of versatile organics and have a wide range of applications from explosives to dyes. Unfortunately, 4-nitrophenol (4NP) and 3-methyl-4-nitrophenol (3M4NP) are also common industrial pollutants that pose environmental and health risks. 4-Methyl-5-nitrocatechol (4M5NC) oxygenase (encoded by the dntB gene) isolated from *Burkholderia* is an FAD dependent enzyme that can catalyze the NADPH assisted oxidation of 4M5NC to 2-hydroxy-5-methylquinone. Leungsakul *et al.* constructed error-prone libraries of the 4M5NC oxygenase to screen for variants capable of oxidizing 4NP and 3M4NP (87). The entire library was screened on plates using the nylon membrane agar colony-screening method. The 4M5NC variant Met22Leu/Leu380Ile was identified after overnight incubation at 37C on 4NP plates. This variant showed a 4-10 fold higher rate of nitrite release during the oxidation of 4NP and 3M4NP relative to the WT enzyme. Also, the variant had ~11-fold increase in catalytic activity towards 4NP oxidation. In addition, site saturation mutagenesis at residues 22 and 380 indicated the synergistic nature of the Leu22 and Ile380 mutations.

Another pollutant, s-triazine, can be hydrolyzed by two distinct enzymes AtzA and TriA by dechlorination and deamination respectively. Although these two enzymes have little substrate overlap and catalyze different bond hydrolysis they differ by only nine amino acids. Raillard *et al.* used DNA shuffling to create a 1600 member library of chimeric enzymes and screened them against a synthetic library of 15 triazines using mass spectrometry (88). A panel of chimeric enzymes was isolated that recognized substrates from the synthetic library not recognized by either parental enzyme.

Alcade *et al.* have developed colorimetric assays for the laccases catalyzed biodegradation of polycyclic hydrocarbons (89). Laccases are fungal enzymes that can facilitate the oxidation of phenolic compounds to the corresponding quinones. In order to monitor the oxidation of anthracene by laccases to 9,10-anthraquinone, they reduced the 9,10-anthraquinone to 9,10-anthrahydroquinone (Abs. Max. 419nm, ε = 4000 $M^{-1}cm^{-1}$) using a water-soluble solution of sodium borohydride. The activity of laccase towards anthracene could thus be monitored spectrophotometrically.

Alcalde *et al.* have reviewed the evolution and application of bioremediation enzymes (90).

1.4.6 Isomerases & Ligases

Isomerases catalyze the inter-conversion of isomers. This class of enzymes includes racemases, epimerases, mutases and cis-trans isomerases. Ligases are enzymes that catalyze the coupling of two molecules using a high-energy phosphate source like ATP.

This class of enzymes includes the acyl-CoA ligases that catalyze the formation of a C-S bond, synthases like peptide synthase that aid C-N bond formation, carboxylases that catalyze C-C bond joining, DNA & RNA ligases that ligate nucleic acids, chelatases that catalyze the reaction between metals and tetrapyrroles and lastly aminoacyl-tRNA synthetases that are essential for protein synthesis. From a directed evolution standpoint, most effort has been directed towards altering the specificity of aminoacyl tRNA synthetases to incorporate non-natural amino acids.

N'-[(5'-(Phosphoribosyl)formimino]-5aminomimidazole-4-carboxamide ribonucleotide (ProFAR) isomerase (HisA) and phosphoribosylanthranilate (PRA) isomerase are a pair of similar enzymes that are believed to have evolved from a common ancestral enzyme and are involved in the biosynthesis of histidine and tryptophan respectively. These are enzymes of the $(\beta\alpha)_8$ barrel structural family, and since they catalyze similar reactions, Jurgens *et al.* used directed evolution to evolve TrpF like activity in the HisA parent (91). A library of mutants was constructed using DNA shuffling on the *Thermotoga maritime* HisA (tHisA) gene. The plasmid library was then transformed into an *E. coli* dtrpF strain (JMB9) and tHisA variants that could complement tryptophan auxotrophy were selected. Selection yielded three colonies, two of which where identical. *In vitro* steady state kinetic characterization of the two enzyme variants revealed that both enzymes behaved as extremely poor catalysts in the HisA isomerization reaction. Further, the tHisA_2 mutant showed appreciable kinetics with the TrpF reaction and although the catalytic

efficiency of the evolved enzyme was still about 4 orders of magnitude lower than the WT tTrpF, it was deemed primarily a K_M effect.

On the same concept of the patchwork hypothesis (divergent evolution) of enzyme evolution (92), Poelarends *et al.* have studied the evolution of enzymic activity in the tautomerase superfamily (93). Since 4-oxolocrotonate tautomerase (4-OT) and trans-3-chloroacrylic acid dehalogenase (CaaD) are members of the same tautomerase family but catalyze completely different reaction chemistries, they mutated the two active-site residues in 4-OT to the corresponding amino acids in CaaD. The double mutant Leu8Arg/Ile52Glu showed a ~32-fold improvement in catalytic efficiency for the catalysis of the dehalogenation reaction.

Alanine racemases are bacterial enzymes that catalyze the interconversion of L & D alanine. Pyridoxal 5'-phospate (PLP) is an essential cofactor for a wide array of enzymic reactions like aldol condensations, transaminations and racemizations. The alanine racemase (Alr) from *Geobacillus stearothermophilus* is a PLP dependent enzyme. The L-threonine aldolase from *Thermatoga maritime* is an evolutionarily unrelated PLP dependent enzyme. Although the two enzymes have completely different reaction chemistries and act on different substrates, the first step in the mechanism of both enzymes is the formation of aldimine intermediates between PLP and their respective substrates. Seebeck and Hilvert have taken advantage of this common link to engineer aldolase activity into Alr (94). Since the Tyr265 of Alr is the base that catalyzes the deprotonation step during racemization, replacement of this tyrosine with alanine served

two functions. First, since the Tyr265 is essential for racemization activity, the 265Ala mutant had a much lower racemase activity. Second, the Tyr265Ala substitution decreased sterics and increased the volume of the binding pocket to allow for binding β-Phenylserine and initiate the reaction with PLP to form the aldimine. The AlrTyr265Ala variant showed a 2.3 x 10^5 fold increase in aldolase activity and a $4x10^3$ fold decrease in racemase activity; a remarkable change in reaction preference based on a single point mutation.

A series of papers from the Hilvert group describes the evolution of chorismate mutase for the optimization of activity (95, 96)and re-engineering the enzyme topology (97, 98).

In WT *E. coli* strains, DsbA and DsbC are primarily responsible for disulfide bond formation in the periplasm. Although, these structurally homologous proteins contain a CXXC thioredoxin (Trx) active-site motif they catalyze complimentary functions, DsbA functions as an oxidase and DsbC functions as an isomerase. Segatori *et al.* engineered a set of DsbC-DsbA and a DsbC-TrxA chimeric proteins that are capable of catalyzing both protein oxidation and disulfide isomerization in the periplasm (99). The engineered chimeras could complement both Dsbc activity, as measured by a tPA refolding assay, and DsbA activity, as confirmed using alkaline phosphatase and cell motility assays.

Since the cell motility is critically dependent on the oxidation state of the periplasm, Masip *et al.* have engineered a disulfide bond formation pathway independent of the DsbA-DsbB catalytic system by using motility assays (100). By constructing a double-site saturation library of the central XX residues in the CXXC TrxA motif, they isolated

variants that could complement DsbA-DsbB knockouts in restoring cell motility. Interestingly, the TrxA(CACC) and TrxA(CACA) variants behaved like ferredoxin type proteins by associating with [2Fe-2S] clusters (Figure 1.11). The ease of conversion of a reductase into an oxidase, based on just two mutations, indicates the adaptability of these essential proteins.

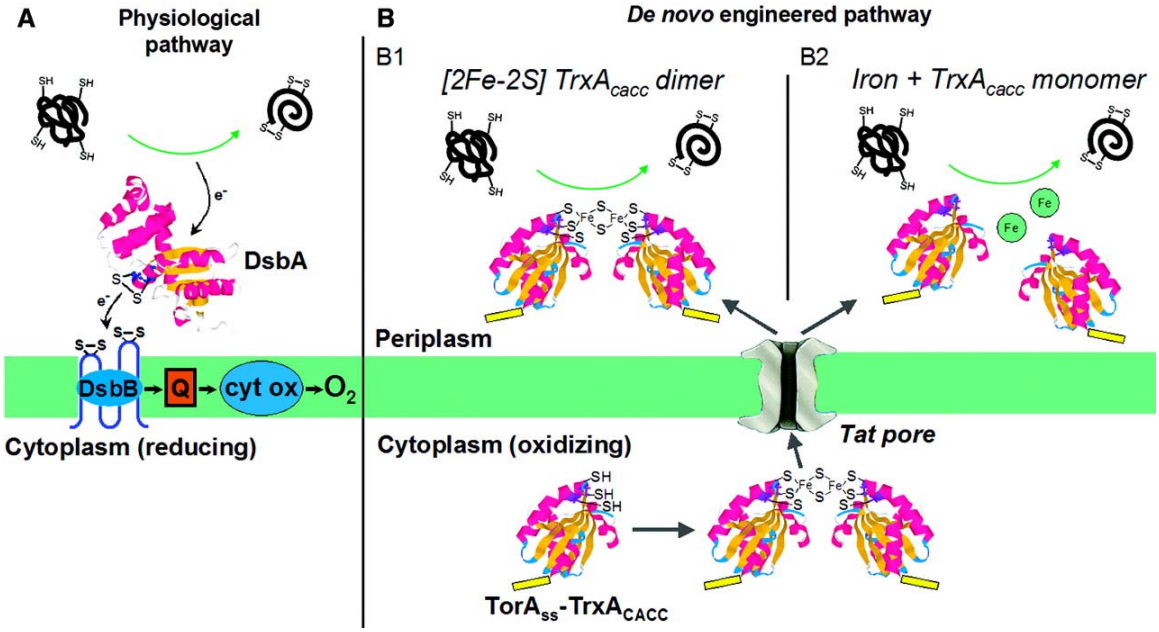


Figure 1.11: "Physiological and engineered disulfide bond formation pathways in the *E. coli* periplasm. (**A**) In the physiological pathway, DsbB provides oxidizing equivalents to DsbA, which acts as a general catalyst of protein thiol oxidation. e^-, electrons; Q, quinones; cyt ox, cytochrome oxidases. (**B**) Pathway for the formation of disulfide bonds by ssTorA-TrxA(CACC). Assembly of the [2Fe-2S] cluster occurs in strains with oxidizing cytoplasm (*trxB* or *trxB gor AhpC** mutants), and the dimeric form of the protein is subsequently exported through the Tat translocator. B1: Catalysis by dimeric TrxA(CACC) with the [2Fe-2S] center intact. B2: Catalysis by a two-part system in which the TrxA(CACC) dimer dissociates in the periplasm to release free iron and thioredoxin. Stoichiometric transfer of a disulfide from the cytoplasm may also be occurring (not shown)." Figure and caption from Masip *et al.* Science (2004), 303, 1185. Reprinted with permission from AAAS.

Aminoacyl-tRNA-synthetases are enzymes responsible for charging (ligating) the tRNA with its cognate amino acid. Given the evolutionary constraints on the responsibility of these enzymes to ensure the fidelity of cellular protein synthesis it would seem likely that these enzymes would have extremely narrow substrate specificities. Although non-discriminating tRNA-synthetases have been identified and characterized (101), by and large, these are highly specific enzymes in the context of their natural substrates. The substrate-binding pocket of aminoacyl-tRNA-synthetases, however, has been shown to be surprisingly plastic (102) with respect to directed evolution of new substrate specifcities and protein engineers have taken advantage of this plasticity to incorporate an extensive array of amino acids into recombinant proteins.

Link *et al.* have reported a general flow-cytometric assay to screen for variants of the *E. coli* methionyl-tRNA-synthetase (MetRS) that can selectively incorporate reactive amino acids (103). A methionine auxotroph, M15MA, was transformed with a multiple-residue saturation library of the MetRS and the cells grown in minimal medium with the 19 natural amino acids and azidonorleucine (ANL). Simultaneously, OmpC expression was induced by the addition of IPTG. To select for cells expressing ANL incorporated recombinant OmpC, they were tagged sequentially with biotin-PEO-cyclooctene and fluorescent avidin (Figure 1.12). Fluorescent cells were then sorted using a flow-cytometer. Analysis of three of the variants isolated indicated that the Leu13Gly change was essential to create more room in the binding pocket of the MetRS to accommodate ANL. Construction of MetRSLeu13Gly variant confirmed that the point mutation

conferred the ability to incorporate ANL (>95%) into recombinant proteins like Dihydrofolate reductase (DHFR).

The Schultz group has developed a general scheme based on the orthogonal amber suppressor *Methanococcus jannaschii* tyrosyl-tRNA (MjtRNATyr) and the alteration of the substrate specificity of the cognate tyrosyl-tRNA synthetase(MjtTyrRS) in *E. coli*. A random library of MjTyrRS variants is constructed and subjected to a two-step positive and negative selection screen. In the first step, MjTyrRS mutants that can successfully charge MjtRNA are selected for by incorporating an amber codon into the chloramphenicol acteyl transferase gene. Survival on chloramphenicol plates ensures read through suppression. In the second step, variants isolated based on the first screen were put through a negative selection based on barnase. Amber mutations were introduced at permissive sites in the toxic barnase gene and the cells grown in the absence of the non-natural amino acid. If the synthetase variant could incorporate a natural amino acid at the permissive site, the cells would die and would thus be selected against. Using this strategy, they have identified MjTyrRS variants that incorporate more than 30 different amino acids in proteins in response to the amber codon (104).

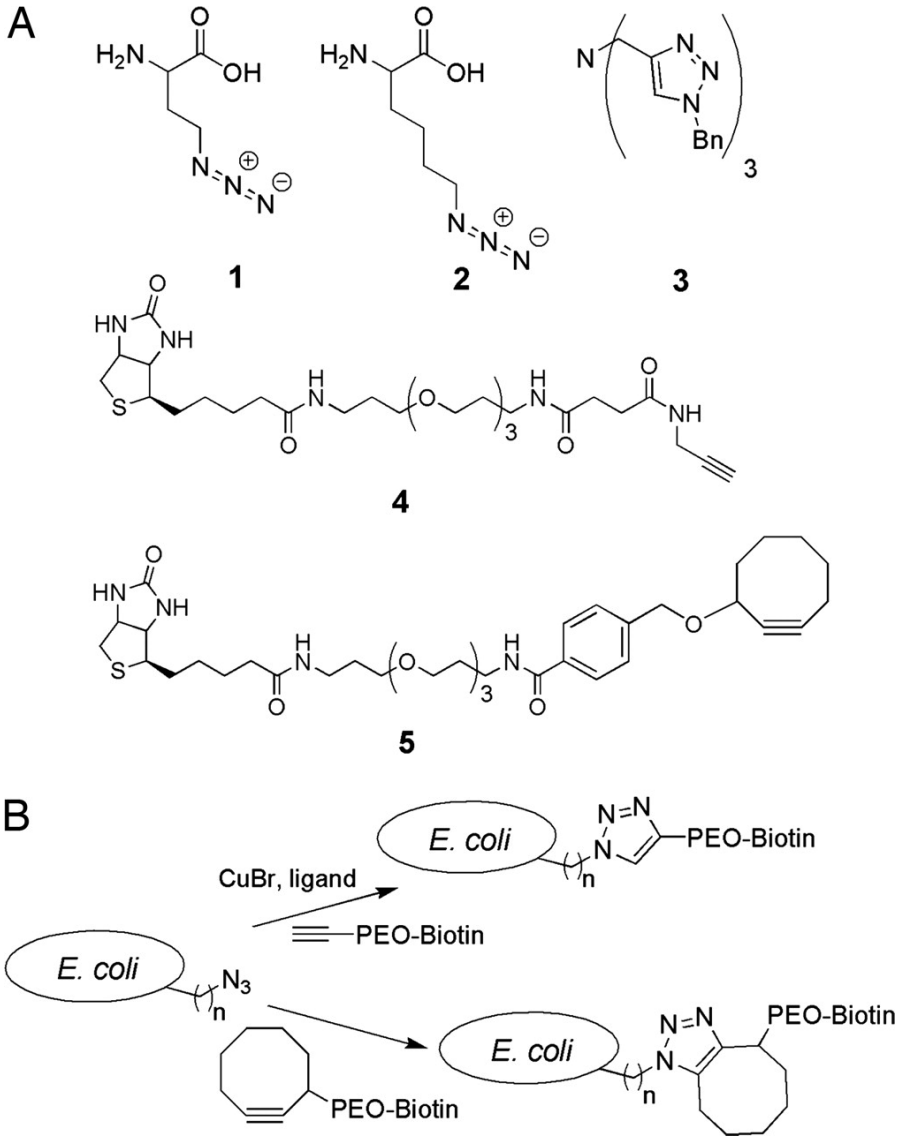


Figure 1.12: "Noncanonical amino acids and tagging reagents. (*A*) **1**, AHA; **2**, ANL; **3**, Tris(triazolyl)amine ligand for copper-catalyzed azide–alkyne ligation; **4**, biotin–PEO–propargylamide; **5**, biotin–PEO–cyclooctyne. (*B*) Scheme for biotin tagging of azide-functionalized *E. coli* cell surfaces. **1** or **2** is incorporated into OmpC to display the azide on the cell surface. Cell surface azides react either by Cu-catalyzed azide–alkyne ligation (top route) or by strain-promoted azide–alkyne ligation (bottom route)." Figure and caption reproduced from Link *et al.* Proc Natl Acad Sci USA (2006), 103, 10180. Copyright (c) 2006 by the National Academy of Sciences.

Parental Enzyme(s)	WT substrate/ New substrate. Specific for new substrate (Y/N)	WT enzyme k_{cat}/K_M ($M^{-1}s^{-1}$) (pref. substrate)	Best Enzyme variant k_{cat}/K_M ($M^{-1}s^{-1}$)	Comments/New activity present in parent enzyme (Y/N)	Ref
Rational redesign					
I-MsoI	**G**GTCGTGAGAC**C**/ **C**GTCGTGAGAC**A** (Y)	--	--	~7700-fold improvement in affinity (Y)	(17)
γ-Humulene synthase	Humulene/ Longifolene (Y)	$5.1* 10^3$	$1.8 * 10^4$	Set of enzymes that yield different sesquiterpenes	(18)
Restriction endonucleases					
I-CreI, I-DmoI	GCCTTGCCGGGTAAGTTCCGGCGCG ; **CAAAACGTCGT**GAGACAGTTTGGT /	--	--	Similar k_{cat}s; not as active as WT enzyme (N)	(24)
NotI	GC**G**CCGC/ GC**T**GCCGC (N)	--	--	Relaxed specificity	(25)
BsoBI	C**Py**CG**Pu**G/ C**C**CG**G**G (Y)	--	--	~70-fold in affinity; not as active as WT enzyme (Y)	(26)
I-SceI	**T**AGGGATAACAGGGTAAT/ **C**AGGGATAACAGGGTAAT (Y)	--	--	~12-fold specificity shift; no improvement in affinity (Y)	(28)

Parental Enzyme(s)	WT substrate/ New substrate. Specific for new substrate (Y/N)	WT enzyme k_{cat}/K_M ($M^{-1}s^{-1}$) (pref. substrate)	Best Enzyme variant k_{cat}/K_M ($M^{-1}s^{-1}$)	Comments/New activity present in parent enzyme (Y/N)	Ref
Recombinases					
Cre	ATA**AC**T**T**CGTATA / ATA**TA**T**A**CGTATA (Y)	--	--	~9000-fold change in specificity (N)	(33)
Cre	ATAACT**TC**GTATA / ATAACT**CT**GTATA (Y)	--	--	Similar rates, 8-fold lower affinity (N)	(34)
Enzymes for Organic synthesis					
NanA (Aldolase)	D-Sialic Acid / L-3-deoxy-manno-2-octulosonic acid (Y)	$4.0*10^3$	$3.7*10^3$	Active and selective enzyme (Y)	(42)
AgaY (Aldolase)	Fructose 1,6-bisphosphate/ Tagose 1,6-bisphosphate # (Y)	$6.2*10^4$	$1.3*10^4$	The enzyme is only mildly specific compared to WT (Y)	(44)
GalK	D-galactose/ L-glucose (N)	$8.6*10^2$	$4.0*10^2$	Anomeric enzyme that still prefers the WT substrate (Y)	(48)

Parental Enzyme(s)	WT substrate/ New substrate. Specific for new substrate (Y/N)	WT enzyme k_{cat}/K_M ($M^{-1}s^{-1}$) (pref. substrate)	Best Enzyme variant k_{cat}/K_M ($M^{-1}s^{-1}$)	Comments/New activity present in parent enzyme (Y/N)	Ref
Enzymes for organic synthesis					
CHMO	R-lactone/ S-lactone[#] (Y)	--	--	Change in ee; no information on activity (Y)	(49)
Acylases and Proteases					
SY-77 acylase	Gluataryl-7-aminocephalosporanic acid / adipyl-7-aminodesacetoxycephalosporanic acid (N)	$1.3*10^5$	$3.9*10^3$	8-fold improvement wrt ADCA (Y)	(55)
Trypsin	Lys(Arg) / Phe(Tyr) (Y)	$1.2*10^6$	$2.0*10^4$	Specificity similar to chymotrypsin (Y)	(58)
HIV-1 Protease	p6/B-Gal/TNFα-BGAL (Y)	--	--	Very little activity with the WT substrate (N)	(64)
OmpT	Arg↓Arg/Arg↓Val (N)	$1.4*10^3$	$1.7*10^5$	60-fold improvement in Arg-Val activity (Y)	(66)

Parental Enzyme(s)	WT substrate/ New substrate. Specific for new substrate (Y/N)	WT enzyme k_{cat}/K_M ($M^{-1}s^{-1}$) (pref. substrate)	Best Enzyme variant k_{cat}/K_M ($M^{-1}s^{-1}$)	Comments/New activity present in parent enzyme (Y/N)	Ref
Acylases and Proteases					
D-amino-peptidase	D-Ala-p-nitroanilide/ N^{α},N^{α}-Ac2 -L-Lys-D-Ala-D-Ala (Y)	$1.3*10^6$	$1.3*10^1$	30-fold selectivity for carboxypeptidase activity (Y)	(67)
Trypsin, FactorXa	MOC-D-Nle-Gly-Arg-pNA; Tos-Gly-Pro-Lys-pNA / MS-D-Phe-Gly-Arg-pNA (N)	$7.3*10^6$, $1.1*10^6$	$2.9*10^6$	Chimeric enzyme with a new preferred substrate (Y)	(68)
Transferases					
RatGSTT-2, HumanGST T1-1	7-amino-4-chloromethyl coumarin/ glutathione (Y)	$1.1*10^4$	$1.0*10^4$	Chimeric enzyme with reactivity towards CMAC (Y)	(69)
AspAT	Aspartate / 2-oxovaline (N)	$2.3*10^7$	$7.4*10^3$	~10^5 fold increase in activity with 2-oxovaline	(72)
Aspartate aminotrasnferase	Aspartate/ Tyrosine (Y)	$6.5*10^5$	$1.7*10^5$	300-fold with tyrosine as the substrate (Y)	(74)

Parental Enzyme(s)	WT substrate/ New substrate. Specific for new substrate (Y/N)	WT enzyme k_{cat}/K_M ($M^{-1}s^{-1}$) (pref. substrate)	Best Enzyme variant k_{cat}/K_M ($M^{-1}s^{-1}$)	Comments/New activity present in parent enzyme (Y/N)	Ref
Transferases					
CstII sialyl transferase	Lactose/BODIPY-3SH-lactose (Y)	--	0.6	~150-400-fold increase in the ability to use BODIPY-sugars (N)	(77)
β-glycosidase	o-NP-Galactose/ o-NP-Gal(1→3)Gal (Y)	$4.8*10^4$	--	Selectivity is designed by suppression of hydrolysis (Y)	(80)
β-gus*	β-D-glucuronoside/ β-galactoside (Y)	$\sim 6*10^7$	$\sim 1.0*10^4$	500-fold increase in β–galactoside activity (Y)	(82)
Bioremediation enzymes					
PON1-Human	Paraxon/DEPCyC (Y)	$2.4*10^6$	$3.6*10^5$	~30- fold increase in DEPCyC hydrolysis (Y)	(84)
DhaA	1,2-dibromomethane/1,2,3-Trichloropropane (N)	$3.8*10^3$	$2.8*10^2$	8-fold improvement with TCP (Y)	(85)

Parental Enzyme(s)	WT substrate/ New substrate. Specific for new substrate (Y/N)	WT enzyme k_{cat}/K_M ($M^{-1}s^{-1}$) (pref. substrate)	Best Enzyme variant k_{cat}/K_M ($M^{-1}s^{-1}$)	Comments/New activity present in parent enzyme (Y/N)	Ref
Bioremediation enzymes					
DntB oxygenase	4_methyl-5-nitrocatechol/ 4-nitrophenol (N)	--	$4.5*10^2$	~11-fold increase in rate of 4NP oxidation (Y)	(87)
AtzA, TriA	Atrazine; Aminoatrazine/ Aminopropazine (N)	--	--	New activity not recognized by either parent. (N)	(88)
Isomerases					
HisA	*N*9-[(59-phosphoribosyl)formimino]-5-aminoimidazole-4-carboxamide ribonucleotide/ phosphoribosylanthranilate (Y)	$1.2*10^6$	$1.2*10^2$	Lower k_{cat}/K_M primarily due to high K_M (Y)	(91)
4-OT	2-oxo-4-hexenedioate/ *trans*-3-chloroacrylate (N)	$1.9*10^7$	$3.5*10^{-1}$	~32-fold increase in efficiency of the dehalogenation reaction (Y)	(93)
Alr Alanine racemase	Alanine/ β-Phenylserine (No information)	--	$1.1*10^1$	$2.3*10^5$ fold increase in aldolase activity and $4*10^3$ fold decrease in racemase activity. No information on overall specificity. (Y)	(94)

Parental Enzyme(s)	WT substrate/ New substrate. Specific for new substrate (Y/N)	WT enzyme k_{cat}/K_M ($M^{-1}s^{-1}$) (pref. substrate)	Best Enzyme variant k_{cat}/K_M ($M^{-1}s^{-1}$)	Comments/New activity present in parent enzyme (Y/N)	Ref
Ligases					
MetRS Methionyl-tRNA-synthetase	Methionine/ Azidonorleucine (Y)	$5.5*10^5$	$1.6*10^3$	Although in vitro characterization indicated a non-specific enzyme, in vivo assays confirmed selectivity. (Y)	(103)
MjtTyrRS Tyrosyl-tRNA-synthetase	Tyrosine/ L-homoglutamine (Y)	--	--	Set of enzymes that selectively incorporate non-natural amino acids. (N)	(104)

Table 1.1: Summary of activity and specificity of enzymes engineered by directed evolution highlighted in this chapter.

#: Product is shown instead of the substrate

1.5 Conclusions

It is clear that the utility of modification of enzyme property through directed evolution has been demonstrated across all families of enzymes, with diverse reaction chemistries. In spite of this, however, developing high-throughput assays for screening large libraries ($>10^7$ members) remains a formidable challenge. This is especially true for enzymes not amenable to genetic selections. The ability to screen large libraries to sift effectively through sequence space is particularly important in the context of recognizing novel substrates not recognized by the parent enzyme. Yet another and often ignored challenge in the alteration of substrate selectivity is the ability engineer enzymes with selectivites and activities comparable to the natural enzymes. Indeed, there are very few examples of enzymes evolved *in vitro* that have catalytic efficiencies and substrate profiles that are comparable to their natural counterparts (Table 1.1). Although enzymes with higher efficiency are almost always desirable, cross-reactivity and specificity depend on the actual enzyme and application.

1.6 References:

1. Schmid, A., Dordick, J. S., Hauer, B., Kiener, A., Wubbolts, M. & Witholt, B. (2001) *Nature* **409,** 258-68.
2. Schoemaker, H. E., Mink, D. & Wubbolts, M. G. (2003) *Science* **299,** 1694-7.
3. Gupta, M. N. & Roy, I. (2004) *Eur J Biochem* **271,** 2575-83.
4. Kazlauskas, R. J. (2005) *Curr Opin Chem Biol* **9,** 195-201.
5. Wang, J. D., Herman, C., Tipton, K. A., Gross, C. A. & Weissman, J. S. (2002) *Cell* **111,** 1027-39.
6. Orencia, M. C., Yoon, J. S., Ness, J. E., Stemmer, W. P. & Stevens, R. C. (2001) *Nat Struct Biol* **8,** 238-42.
7. Weinreich, D. M., Delaney, N. F., Depristo, M. A. & Hartl, D. L. (2006) *Science* **312,** 111-4.
8. Bolon, D. N., Voigt, C. A. & Mayo, S. L. (2002) *Curr Opin Chem Biol* **6,** 125-9.
9. Park, H. S., Nam, S. H., Lee, J. K., Yoon, C. N., Mannervik, B., Benkovic, S. J. & Kim, H. S. (2006) *Science* **311,** 535-8.
10. Dwyer, M. A., Looger, L. L. & Hellinga, H. W. (2004) *Science* **304,** 1967-71.
11. Ueno, T., Yokoi, N., Unno, M., Matsui, T., Tokita, Y., Yamada, M., Ikeda-Saito, M., Nakajima, H. & Watanabe, Y. (2006) *Proc Natl Acad Sci U S A* **103,** 9416-21.
12. Watanabe, Y. (2002) *Curr Opin Chem Biol* **6,** 208-16.

13. Bradley, L. H., Wei, Y., Thumfort, P., Wurth, C. & Hecht, M. H. (2007) *Methods Mol Biol* **352,** 155-66.

14. Wei, Y. & Hecht, M. H. (2004) *Protein Eng Des Sel* **17,** 67-75.

15. Lerner, R. A., Benkovic, S. J. & Schultz, P. G. (1991) *Science* **252,** 659-67.

16. Kortemme, T., Joachimiak, L. A., Bullock, A. N., Schuler, A. D., Stoddard, B. L. & Baker, D. (2004) *Nat Struct Mol Biol* **11,** 371-9.

17. Ashworth, J., Havranek, J. J., Duarte, C. M., Sussman, D., Monnat, R. J., Jr., Stoddard, B. L. & Baker, D. (2006) *Nature* **441,** 656-9.

18. Yoshikuni, Y., Ferrin, T. E. & Keasling, J. D. (2006) *Nature* **440,** 1078-82.

19. Shifman, J. M. & Mayo, S. L. (2003) *Proc Natl Acad Sci U S A* **100,** 13274-9.

20. Kaur, J. & Sharma, R. (2006) *Crit Rev Biotechnol* **26,** 165-99.

21. Chevalier, B. S. & Stoddard, B. L. (2001) *Nucleic Acids Res* **29,** 3757-74.

22. Pingoud, A. & Jeltsch, A. (2001) *Nucleic Acids Res* **29,** 3705-27.

23. Arnould, S., Chames, P., Perez, C., Lacroix, E., Duclert, A., Epinat, J. C., Stricher, F., Petit, A. S., Patin, A., Guillier, S., Rolland, S., Prieto, J., Blanco, F. J., Bravo, J., Montoya, G., Serrano, L., Duchateau, P. & Paques, F. (2006) *J. Mol. Biol.* **355,** 443-58.

24. Chevalier, B. S., Kortemme, T., Chadsey, M. S., Baker, D., Monnat, R. J. & Stoddard, B. L. (2002) *Mol Cell* **10,** 895-905.

25. Samuelson, J. C., Morgan, R. D., Benner, J. S., Claus, T. E., Packard, S. L. & Xu, S. Y. (2006) *Nucleic Acids Res* **34,** 796-805.

26. Zhu, Z., Zhou, J., Friedman, A. M. & Xu, S. Y. (2003) *J Mol Biol* **330,** 359-72.

27. Chen, Z. & Zhao, H. (2005) *Nucleic Acids Res* **33,** e154.

28. Doyon, J. B., Pattanayak, V., Meyer, C. B. & Liu, D. R. (2006) *J Am Chem Soc* **128,** 2477-84.

29. Dhanasekaran, M., Negi, S. & Sugiura, Y. (2006) *Acc Chem Res* **39,** 45-52.

30. Kim, Y. G., Cha, J. & Chandrasegaran, S. (1996) *Proc Natl Acad Sci U S A* **93,** 1156-60.

31. Urnov, F. D., Miller, J. C., Lee, Y. L., Beausejour, C. M., Rock, J. M., Augustus, S., Jamieson, A. C., Porteus, M. H., Gregory, P. D. & Holmes, M. C. (2005) *Nature* **435,** 646-51.

32. Grindley, N. D., Whiteson, K. L. & Rice, P. A. (2006) *Annu Rev Biochem* **75,** 567-605.

33. Buchholz, F. & Stewart, A. F. (2001) *Nat Biotechnol* **19,** 1047-52.

34. Santoro, S. W. & Schultz, P. G. (2002) *Proc Natl Acad Sci U S A* **99,** 4185-90.

35. Jaeger, K. E. & Eggert, T. (2004) *Curr Opin Biotechnol* **15,** 305-13.

36. Westlake, A. C., Harford-Cross, C. F., Donovan, J. & Wong, L. L. (1999) *Eur J Biochem* **265,** 929-35.

37. Glieder, A., Farinas, E. T. & Arnold, F. H. (2002) *Nat Biotechnol* **20,** 1135-9.

38. Peters, M. W., Meinhold, P., Glieder, A. & Arnold, F. H. (2003) *J Am Chem Soc* **125,** 13442-50.

39. Urlacher, V. B., Lutz-Wahl, S. & Schmid, R. D. (2004) *Appl Microbiol Biotechnol* **64,** 317-25.

40. Urlacher, V. B. & Schmid, R. D. (2004) *Methods Enzymol* **388,** 208-24.

41. de Raadt, A. & Griengl, H. (2002) *Curr Opin Biotechnol* **13,** 537-42.

42. Hsu, C. C., Hong, Z., Wada, M., Franke, D. & Wong, C. H. (2005) *Proc Natl Acad Sci U S A* **102,** 9122-6.

43. Stemmer, W. P. (1994) *Nature* **370,** 389-91.

44. Williams, G. J., Domann, S., Nelson, A. & Berry, A. (2003) *Proc Natl Acad Sci U S A* **100,** 3143-8.

45. Samland, A. K. & Sprenger, G. A. (2006) *Appl Microbiol Biotechnol* **71,** 253-64.

46. Franke, D., Hsu, C. C. & Wong, C. H. (2004) *Methods Enzymol* **388,** 224-38.

47. Reetz, M. T. (2004) *Proc Natl Acad Sci U S A* **101,** 5716-22.

48. Hoffmeister, D., Yang, J., Liu, L. & Thorson, J. S. (2003) *Proc Natl Acad Sci U S A* **100,** 13184-9.

49. Reetz, M. T., Brunner, B., Schneider, T., Schulz, F., Clouthier, C. M. & Kayser, M. M. (2004) *Angew Chem Int Ed Engl* **43,** 4075-8.

50. Reetz, M. T., Daligault, F., Brunner, B., Hinrichs, H. & Deege, A. (2004) *Angew Chem Int Ed Engl* **43,** 4078-81.

51. Schmidt, M. & Bornscheuer, U. T. (2005) *Biomol Eng* **22,** 51-6.

52. Nicholas, K. M., Wentworth, P., Jr., Harwig, C. W., Wentworth, A. D., Shafton, A. & Janda, K. D. (2002) *Proc Natl Acad Sci U S A* **99,** 2648-53.

53. Ueno, T., Suzuki, M., Goto, T., Matsumoto, T., Nagayama, K. & Watanabe, Y. (2004) *Angew Chem Int Ed Engl* **43,** 2527-30.

54. Thomas, C. M. & Ward, T. R. (2005) *Chem Soc Rev* **34,** 337-46.

55. Sio, C. F., Otten, L. G., Cool, R. H. & Quax, W. J. (2003) *Biochem Biophys Res Commun* **312,** 755-60.

56. Otten, L. G., Sio, C. F., Vrielink, J., Cool, R. H. & Quax, W. J. (2002) *J Biol Chem* **277,** 42121-7.

57. Sio, C. F. & Quax, W. J. (2004) *Curr Opin Biotechnol* **15,** 349-55.

58. Hedstrom, L., Szilagyi, L. & Rutter, W. J. (1992) *Science* **255,** 1249-53.

59. Gupta, R., Beg, Q. K. & Lorenz, P. (2002) *Appl Microbiol Biotechnol* **59,** 15-32.

60. Maurer, K. H. (2004) *Curr Opin Biotechnol* **15,** 330-4.

61. Inouye, M., Fu, X. & Shinde, U. (2001) *Nat Struct Biol* **8,** 321-5.

62. Takagi, H. & Takahashi, M. (2003) *Appl Microbiol Biotechnol* **63,** 1-9.

63. Bryan, P. N. (2000) *Biochim Biophys Acta* **1543,** 203-222.

64. O'Loughlin, T. L., Greene, D. N. & Matsumura, I. (2006) *Mol Biol Evol* **23,** 764-72.

65. McCarter, J. D., Stephens, D., Shoemaker, K., Rosenberg, S., Kirsch, J. F. & Georgiou, G. (2004) *J Bacteriol* **186,** 5919-25.

66. Olsen, M. J., Stephens, D., Griffiths, D., Daugherty, P., Georgiou, G. & Iverson, B. L. (2000) *Nat Biotechnol* **18,** 1071-4.

67. Delmarcelle, M., Boursoit, M. C., Filee, P., Baurin, S. L., Frere, J. M. & Joris, B. (2005) *Protein Sci* **14,** 2296-303.

68. Hopfner, K. P., Kopetzki, E., Kresse, G. B., Bode, W., Huber, R. & Engh, R. A. (1998) *Proc Natl Acad Sci U S A* **95,** 9813-8.

69. Griswold, K. E., Kawarasaki, Y., Ghoneim, N., Benkovic, S. J., Iverson, B. L. & Georgiou, G. (2005) *Proc Natl Acad Sci U S A* **102,** 10082-7.

70. Griswold, K. E., Aiyappan, N. S., Iverson, B. L. & Georgiou, G. (2006) *J Mol Biol*.

71. Emren, L. O., Kurtovic, S., Runarsdottir, A., Larsson, A. K. & Mannervik, B. (2006) *Proc Natl Acad Sci U S A* **103,** 10866-70.

72. Yano, T., Oue, S. & Kagamiyama, H. (1998) *Proc Natl Acad Sci U S A* **95,** 5511-5.

73. Onuffer, J. J. & Kirsch, J. F. (1995) *Protein Sci* **4,** 1750-7.

74. Rothman, S. C., Voorhies, M. & Kirsch, J. F. (2004) *Protein Sci* **13,** 763-72.

75. Rothman, S. C. & Kirsch, J. F. (2003) *J Mol Biol* **327,** 593-608.

76. Chow, M. A., McElroy, K. E., Corbett, K. D., Berger, J. M. & Kirsch, J. F. (2004) *Biochemistry* **43,** 12780-7.

77. Aharoni, A., Thieme, K., Chiu, C. P., Buchini, S., Lairson, L. L., Chen, H., Strynadka, N. C., Wakarchuk, W. W. & Withers, S. G. (2006) *Nat Methods* **3,** 609-14.

78. Fernandez-Gacio, A., Uguen, M. & Fastrez, J. (2003) *Trends Biotechnol* **21,** 408-14.

79. Love, K. R., Swoboda, J. G., Noren, C. J. & Walker, S. (2006) *Chembiochem* **7,** 753-6.

80. Feng, H. Y., Drone, J., Hoffmann, L., Tran, V., Tellier, C., Rabiller, C. & Dion, M. (2005) *J Biol Chem* **280,** 37088-97.

81. Hancock, S. M., Vaughan, M. D. & Withers, S. G. (2006) *Curr Opin Chem Biol* **10,** 509-19.

82. Matsumura, I. & Ellington, A. D. (2001) *J Mol Biol* **305,** 331-9.

83. Roodveldt, C., Aharoni, A. & Tawfik, D. S. (2005) *Curr Opin Struct Biol* **15,** 50-6.

84. Aharoni, A., Gaidukov, L., Yagur, S., Toker, L., Silman, I. & Tawfik, D. S. (2004) *Proc Natl Acad Sci U S A* **101,** 482-7.

85. Bosma, T., Damborsky, J., Stucki, G. & Janssen, D. B. (2002) *Appl Environ Microbiol* **68,** 3582-7.

86. Janssen, D. B. (2004) *Curr Opin Chem Biol* **8,** 150-9.

87. Leungsakul, T., Johnson, G. R. & Wood, T. K. (2006) *Appl Environ Microbiol* **72,** 3933-9.

88. Raillard, S., Krebber, A., Chen, Y., Ness, J. E., Bermudez, E., Trinidad, R., Fullem, R., Davis, C., Welch, M., Seffernick, J., Wackett, L. P., Stemmer, W. P. & Minshull, J. (2001) *Chem Biol* **8,** 891-8.

89. Alcalde, M., Farinas, E. T. & Arnold, F. H. (2004) *J Biomol Screen* **9,** 141-6.

90. Alcalde, M., Ferrer, M., Plou, F. J. & Ballesteros, A. (2006) *Trends Biotechnol* **24,** 281-7.

91. Jurgens, C., Strom, A., Wegener, D., Hettwer, S., Wilmanns, M. & Sterner, R. (2000) *Proc Natl Acad Sci U S A* **97,** 9925-30.

92. Rison, S. C. & Thornton, J. M. (2002) *Curr Opin Struct Biol* **12,** 374-82.

93. Poelarends, G. J., Almrud, J. J., Serrano, H., Darty, J. E., Johnson, W. H., Jr., Hackert, M. L. & Whitman, C. P. (2006) *Biochemistry* **45,** 7700-8.

94. Seebeck, F. P. & Hilvert, D. (2003) *J Am Chem Soc* **125,** 10158-9.

95. Vamvaca, K., Butz, M., Walter, K. U., Taylor, S. V. & Hilvert, D. (2005) *Protein Sci* **14,** 2103-14.

96. Taylor, S. V., Walter, K. U., Kast, P. & Hilvert, D. (2001) *Proc Natl Acad Sci U S A* **98,** 10596-601.

97. Vamvaca, K., Vogeli, B., Kast, P., Pervushin, K. & Hilvert, D. (2004) *Proc Natl Acad Sci U S A* **101,** 12860-4.

98. MacBeath, G., Kast, P. & Hilvert, D. (1998) *Science* **279,** 1958-61.

99. Segatori, L., Paukstelis, P. J., Gilbert, H. F. & Georgiou, G. (2004) *Proc Natl Acad Sci U S A* **101,** 10018-23.

100. Masip, L., Pan, J. L., Haldar, S., Penner-Hahn, J. E., DeLisa, M. P., Georgiou, G., Bardwell, J. C. & Collet, J. F. (2004) *Science* **303,** 1185-9.

101. Schulze, J. O., Masoumi, A., Nickel, D., Jahn, M., Jahn, D., Schubert, W. D. & Heinz, D. W. (2006) *J Mol Biol* **361,** 888-97.

102. Turner, J. M., Graziano, J., Spraggon, G. & Schultz, P. G. (2006) *Proc Natl Acad Sci U S A* **103,** 6483-8.

103. Link, A. J., Vink, M. K., Agard, N. J., Prescher, J. A., Bertozzi, C. R. & Tirrell, D. A. (2006) *Proc Natl Acad Sci U S A* **103,** 10180-5.

104. Xie, J. & Schultz, P. G. (2006) *Nat Rev Mol Cell Biol* **7,** 775-82.

Chapter 2

Engineering of Protease Variants Exhibiting High Catalytic Activity and Exquisite Substrate Selectivity

2.1 Introduction

The reprogramming of enzyme catalytic activity and selectivity is a central issue in protein biochemistry and biotechnology. Numerous structure-guided and directed evolution strategies have been employed in search of enzyme variants that exhibit high catalytic rates with poor or inactive substrates of the parental enzyme (1-19) (Chapter 1). As impressive as these successes have been, the engineering of enzymes that exhibit turnover rates and selectivities with new substrates comparable to their natural counterparts has proven quite a challenge, especially when considering those enzymes for which a genetic selection strategy is not possible.

In particular, enzymes engineered through laboratory evolution involving *in vitro* catalytic assays have often been found lacking, either with respect to turnover rates or selectivity, relative to catalyst-substrate pairs isolated from natural sources. As a typical example, an extensive directed evolution program led to the isolation of *E. coli* β-glucuronidase variants with significant β-galactosidase (10) or xylanosidase (11) activities, but nonetheless even the best clones exhibited k_{cat}/K_m values >1,000 times

lower than those of naturally occurring enzymes such as the *E. coli* β-galactosidase or the *T. saccharolyticum* β-xylosidase.

This trend appears to be general. In a recent comprehensive study, Aaron *et al.* (12) demonstrated that the evolution of higher activity towards poor substrates did not impair the parental catalytic activity and therefore the evolved enzymes exhibited greater promiscuity. Enzymes evolved for higher substrate enantioselectivity often exhibit lower specific activities towards their new substrates relative to their respective parental enzymes (13-15). Similarly, the evolution of highly active variants of aspartate aminotransferases capable of accepting branched or aromatic amino acid substrates was accompanied by a relaxation of the substrate selectivity (16,17)

In nature, the evolution of enzymes occurs as the result of positive selective pressure for turnover of physiological substrates, combined with simultaneous negative selective pressure in order to eliminate completely, or at least drastically suppress, deleterious activities. It follows that the engineering of enzymes exhibiting high catalytic activity and substrate selectivity for a particular desired substrate should be similarly accomplished by implementing selection and counter-selection assay schemes in the laboratory. The recent directed evolution of novel tRNA synthetases and recombinases by capitalizing on *in vivo* selections support this expectation (18-19). Unfortunately, many desired enzyme activities are not amenable to *in vivo* selection strategies because cellular growth cannot be linked to the enzyme activity being sought. Consequently, laboratory directed evolution approaches for the isolation of highly selective enzymes

must rely on *in vitro* catalytic assays whereby the activity towards selection and counter-selection substrates is determined sequentially for each library member (9-11,13,14). The successful implementation of the latter approach hinges on satisfying the following requirements: First, assays that afford the proper dynamic range must be designed since mutant proteins are likely to exhibit drastically different k_{cat} and K_m values for the various selection and counter-selection substrates. Second, the isolation of rare, high activity clones requires the screening of large libraries by assaying each clone towards multiple substrates, thus necessitating the availability of suitable high throughput methods.

With these considerations in mind, we have extended our previous approach (20) and developed a new two-pronged strategy in which catalytic activities over a wide dynamic range for both a selection substrate and one or more counter-selection substrates are quantified *simultaneously* at the single cell level, enabling the rapid screening of mutant libraries. This approach relies on fluorescent substrates of different colors that label the surface of *E. coli* cells upon cleavage by a surface-anchored enzyme (20). Enzymes can be displayed on the surface of Gram-negative bacteria by established techniques (21,22) and thus, access to the fluorescent substrate is assured. The net result is that the cell fluorescence profile accurately reflects the catalytic activity and selectivity of the surface displayed enzyme. Multi-parameter flow cytometry is then employed to isolate clones expressing enzymes having a desired fluorescence profile from large libraries.

The *E. coli* endoprotease OmpT and its Omptin homologues play important roles in pathogenicity, are of significance in protein manufacturing, and have been exploited for biotechnology applications (23,24). OmpT has a strong preference for cleavage between two basic residues (Lys and especially Arg) in the P_1 and P_1' positions of the substrate (25-27). We sought to isolate OmpT variants that exhibit (1) hydrolysis of a substrate that is cleaved poorly by the wild-type enzyme, and (2) a low rate of cleavage of dibasic sequences (preferred by the wild-type enzyme) thus conferring (3) a high selectivity for the new cleavage over the one preferred by the parental enzyme (Figure 2.1A), while maintaining (4) a very high level of catalytic activity.

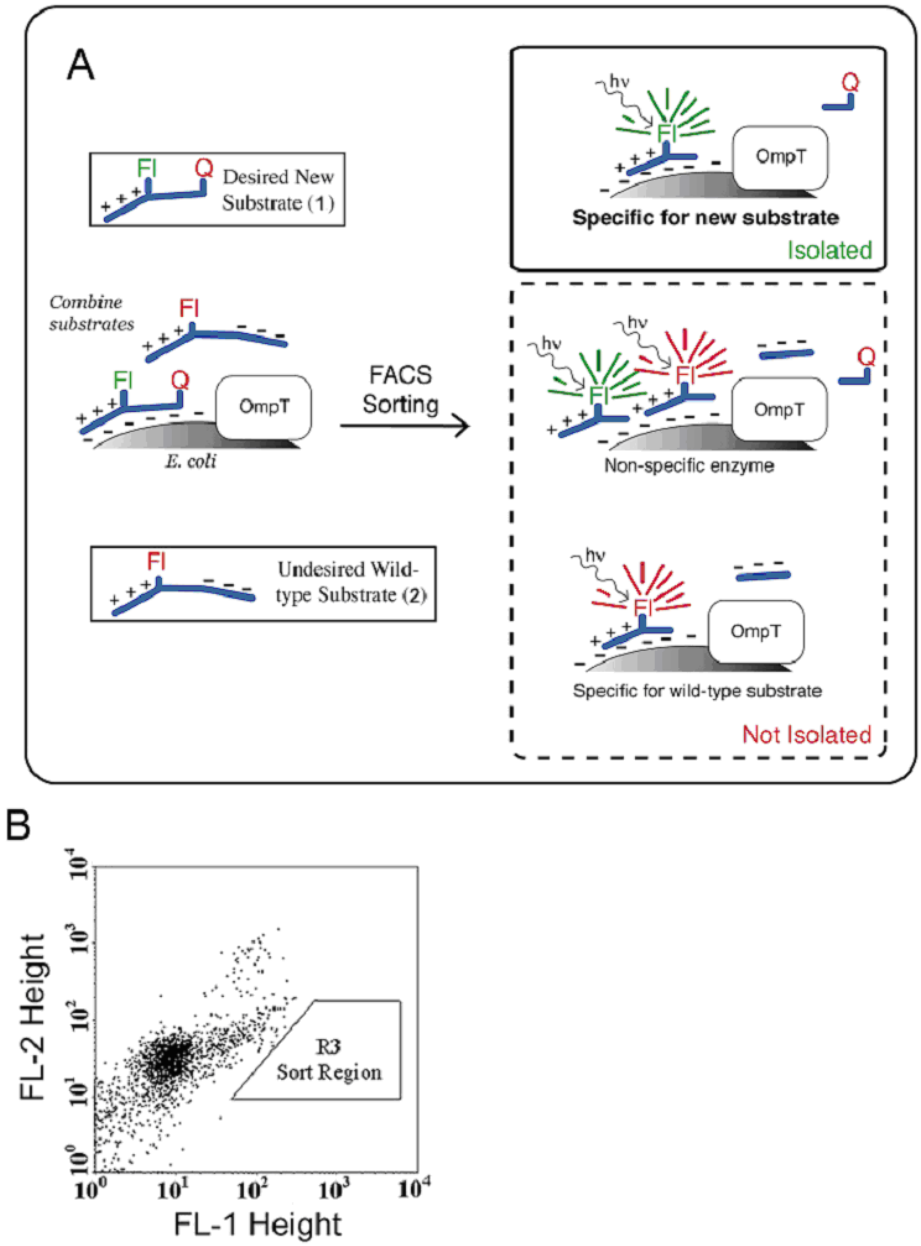


Figure 2.1. (A) Two-color discrimination and selection using the FRET and electrostatic capture substrates; (B) Library sort gate R3, used to isolate positive clones displaying high FL-1 (from the hydrolysis of **1**) and low FL-2 (lack of hydrolysis of **2**) fluorescence.

2.2 Materials and Methods

Substrate synthesis.

The peptides **1a** Ac-CARVGKGRGR-NH_2 and **2a** Ac-EEGGRRIGRGGK-NH_2, were synthesized at the University of Texas Peptide synthesis facility. Tetramethylrhodamine-5-iodoacetamide (TMRIA), 5- carboxytetramethylrhodamine, succinimidyl ester (5-TAMRA, SE) and BODIPY®-FL-SE were purchased from Molecular Probes, Eugene, OR. The peptides **3** WCARVGKGRGR-NH_2 and **4** WEEGGRRIGRGGK-NH_2 (>95% purity) were purchased from Cell Essentials (Boston, MA).

For the synthesis of the FRET substrate **1**, a 50 μL solution the **1a** (5.4 mg, 4.9 μmol) in water was added to 100 μL of 1M Na_2CO_3 and 50 μl of 0.5M $NaHCO_3$. A solution of 50 μL of TMRIA (4.0 mg, 4.8 μmol) in DMF was added to the reaction mixture and stirred at RT for 1h. The reaction mixture was quenched with 5 ml of 0.1% TFA in water and the product purified by FPLC using a 10% -30% acetonitrile gradient. The purified product (2.6 mg, 35% yield) was freeze-dried, dissolved in 500 μL of water and its identity was confirmed by both ^{1}H-NMR and ESI-MS. For the conjugation of BODIPY a 90 μL solution of the purified product (1.9 mg, 1.2 μmol) in water, was mixed with a 90 μL solution of BODIPY®-FL-SE (0.5 mg, 1.3 μmol) in DMF. 25μL of 1M DMAP (25 μmol) was added to the reaction mixture and stirred at RT for 1h. The reaction mixture was quenched with 5ml of 0.1% TFA in water and purified by FPLC

with a 10% -30% acetonitrile gradient. The purified product (0.7 mg, 32% yield) was freeze-dried, dissolved in 500 μL of water and quantified by UV. The product identity was confirmed by both ^{1}H-NMR and ESI-MS.

The electrostatic capture substrate **2** was synthesized exactly as described above for BODIPY conjugation, except that 5-TAMRA, SE was used instead of BODIPY®-FL-SE.

Flow cytometric analysis: *E.coli* UT5600 (*F⁻ ara-14 leuB6 secA6 lacY1 proC14 tsx-67 Δ(ompT-fepC)266 entA403 trpE38 rfbD1 rpsL109(Str^{R}) xyl-5 mtl-1 thi-1*) was transformed with pML19 (28) encoding wild-type OmpT under the control of its native promoter. Overnight cultures of UT5600, UT5600/pML19 were resuspended in 1% sucrose, diluted to 0.01 OD_{600}, labeled for 10 min with 50 nM **1** in 1% sucrose, diluted into 1 ml 1% sucrose, and analyzed using a Becton-Dickinson FACSort.

Molecular Biology methods A library of random mutants was constructed by error-prone PCR using Clonetech's (Palo Alto, CA) Diversify mutagenesis kit. The PCR product, pAMP1 vector DNA (Life Technologies, CA), annealing buffer, and 2 units Uracil DNA glycosylase in a total volume of 20 μl were incubated for 40 minutes at 37^{o}C, followed by 1 hour at 4^{o}C. The reaction mixture was used to electroporate electrocompetent *E. coli* DH10B (F⁻ *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *end*A1 *ara*D139 Δ(*ara, leu*)7697 *gal*U *gal*K λ- *rps*L *nup*G) cells and the

entire library was plated on selective media. The clones were pooled and frozen at –80°C in aliquots.

OmpT mutants in which Ser223 was substituted with Arg, Gly, Leu, Lys, Phe or Trp were constructed using the QuikChange® Site-Directed Mutagenesis Kit (Stratagene, La Jolla CA) with pML19 as the template and appropriate overlapping primer pairs.

Library Screening: Transformants were grown at 37°C in LB media, harvested after 16 hours, washed once with 1% sucrose and resuspended in 1% sucrose. A 50 μL aliquot of the cell suspension in sucrose was added to 949 μl of 1% sucrose and labeled using 1 μl of each **1** & **2** (final concentration 100 nM). A 20 μL aliquot of this labeling reaction was diluted into 1% sucrose and analyzed on the flow-cytometer. Library sorting was performed on a Becton Dickinson FACSCalibur instrument, using gates set based upon FSC/SSC and FL1/FL2 values. A total of ~$3x10^6$ cells were sorted in 29 min and 1050 clones were isolated and then resorted. The collected solution was filtered, and the filters were placed on agar plates containing 200 μg/ml ampicillin. After 14 hours, 99 colonies that grew were individually inoculated into 1 ml LB media containing 200 μg/ml ampicillin. Individual colonies were screened using either substrate **1** or **2** on the flow-cytometer using the exact same protocol described above.

Enzyme purification and kinetic analysis. Proteins were isolated as previously described, with minor modifications (20) to a final purity >90% as determined by SDS-

PAGE. For kinetic analyses, 10-20 nM of the purified enzymes were incubated with 20 μM to 1 mM of the appropriate substrate in 0.1 M Tris, 10 mM EDTA, pH 8.0 at room temperature (25^{o}C) and the reaction was monitored by HPLC on a Phenomenex C18 reverse phase column using the following gradient: 5% AcN/95% H_2O for 1 min, increasing to 95% AcN/5% H_2O over a period of 29 mins and returning to 5% AcN/95% H_2O over 5 mins. The product concentration was determined using the integration areas at 280 nm and the apparent rates were fitted to a Michaelis-Menten by non-linear regression. The cleavage products were determined by LC-MS (ESI) on a Magic 2002 instrument (Micron Bioresources, Auburn, CA).

2.3 Results

Two color Flow Cytometric Screening Strategy: Electrostatic interactions between the negatively charged bacterial surface (29) and substrates **1** and **2** (Figure 2.2) were exploited to capture the fluorescent and positively charged products of the enzymatic cleavage reaction on the surface of the bacteria. The selection substrate **1** is a FRET peptide with a net +3 charge, the minimum charge required for cell surface capture (30). **1** is electrostatically adsorbed on the cell surface but upon cleavage, the *N*-terminal moiety consisting of Ac-NH_2-CA bound to the tetramethylrhodamine (TMR) dye is released from the cell, resulting in accumulation of the *C*-terminal peptide, which is positively charged and contains the BODIPY® (4,4-difluoro-5,7-dimethyl-4-bora-3a,4a-diaza-s-indacene-3-propionic acid) fluorophore. Substrate **1** was used as the selection

substrate because it accumulates on the cell surface prior to cleavage, resulting in a high effective molarity, which in turn, enables the detection of even very weak catalytic activity. The counter-selection substrate **2** was a zwitterionic peptide containing a single fluorophore and at least three positive charges on one side of the scissile bond (Arg-Arg), as well as an equal number of negatively charged groups on the other (Figure 2.1A) (The counter-selection substrate scheme was devised by Jongisk Gam from our group). The intact substrate has no net charge, but enzymatic cleavage generates a positively charged moiety carrying the fluorescent dye that is deposited on the cell surface. Upon incubation with **2**, *ompT*$^{+}$ cells exhibited a >10 fold higher fluorescence relative to the *E. coli ompT*$^{-}$ deletion strain UT5600 (data not shown). The combination of a surface bound selection FRET substrate **1** and zwitterionic counter-selection substrate **2** that is free in solution afforded the proper dynamic range required for isolation of highly active and selective enzymes.

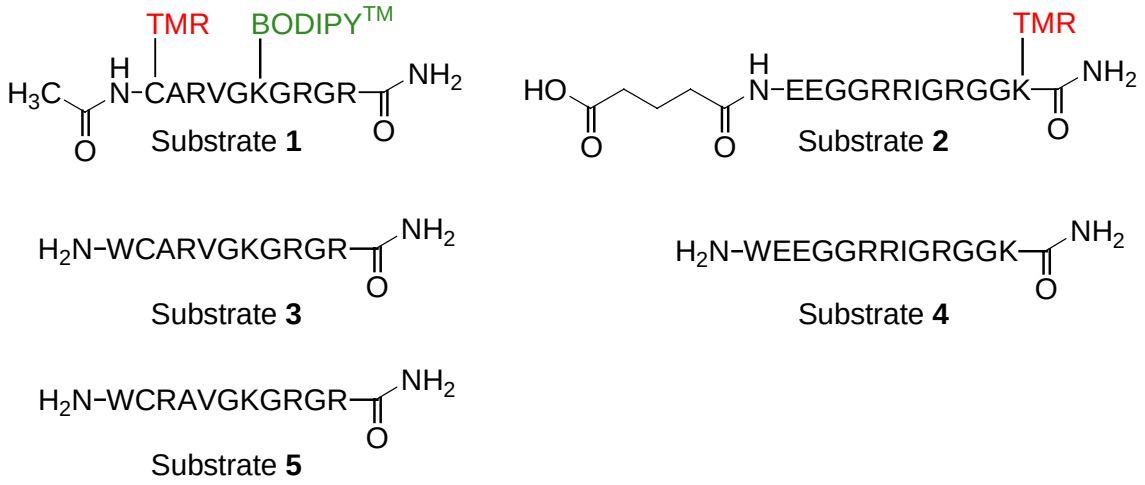


Figure 2.2: Substrates used for the detection of catalytic activity and for the kinetic analysis.

Library Screening: C5 is an OmpT variant exhibiting a 60-fold increase in activity towards cleavage at a non-preferred Arg-Val site (20). However, as is typically the case for engineered enzymes, C5 exhibits relaxed overall specificity, and is not selective for the cleavage of Arg-Val sites. For example, the C5 variant also cleaved peptide substrates at Ala-Arg sequences with a catalytic efficiency even higher than that of the wild-type OmpT (Table 2.1) as well as other sequences (20). The gene encoding C5 downstream from the *ompT* promoter was subjected to random mutagenesis by error-prone PCR (30) and a library of approximately 1 x 10^6 transformants was generated. DNA sequencing of 10 randomly selected clones revealed a mutation rate of 1.1% nucleotides per gene. The *E. coli* library was incubated in a solution of 1% w/v sucrose

(to maintain proper osmotic balance) together with 100 nM of the electrostatic capture substrate **2** and the FRET substrate **1** for 10 minutes. Bacterial cells displaying increased BODIPY® fluorescence (green, emission at 530 nm, FL-1) and reduced TMR fluorescence (red, emission at 560 nm, FL-2) corresponding to high activity with substrate **1** and reduced activity with the counter-selection peptide **2**, respectively, were isolated (Figure 2.1B). A total of 1050 clones were obtained and following resorting, 99 clones were isolated, grown individually in microtiter well plates. Two clones, 1.3.19 and 1.2.19, were selected for further study because they both exhibited high green fluorescence (Figure 2.3) and reduced red fluorescence consistent with the sorting criteria used for their isolation. Sequencing determined that 1.2.19 (Table 2.2) contained Asp208Gly and Asp214Val mutations whereas 1.3.19 contained Ser17Gly and Ser223Arg mutations in addition to those found in C5, the parent enzyme used in this study.

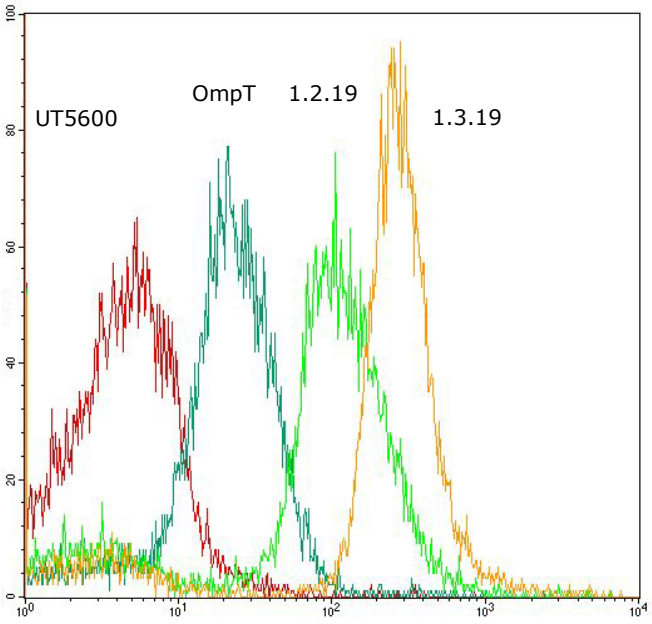


Figure 2.3: Flow-cytometric discrimination of *E. coli* UT5600 (Δ*ompT*) transformed with pML19 (expressing wt OmpT), pML1.2.19 and pML1.3.19 using FRET substrate **1**. Briefly, the cells were washed and resuspended in 1% sucrose and labeled with 50nM (final concentration) of the AR-FRET substrate **1**. A 20ul aliquot of the labeling reaction was transferred to 0.5ml of 1% sucrose and analyzed on the flow-cytometer.

Enzyme	-A-R- [1]			-R-R- [2]			Specificity
	k_{cat} (s^{-1})	K_M (uM)	k_{cat}/K_M ($s^{-1}M^{-1}$)	k_{cat} (s^{-1})	K_M (uM)	k_{cat}/K_M ($s^{-1}M^{-1}$)	-AR-/-RR-
OmpT	3.1±0.5 x10^{-2}	16±5	2.2±0.9 x10^{3}	8.8±0.7	55±9	1.7±0.4 x10^{5}	1.3 x 10^{-2}
C5	1.7±0.5x10^{-2}	1.5± 0.6	1.4±0.8 x10^{4}	0.7± 0.1	2.2 ±0.4	3±1.0 x10^{5}	3.6x 10^{-2}
1.2.19	1.7±0.3	9±1	2.1±0.6 x10^{5}	0.4±0.1	260±90	2±1 x10^{3}	1.0x 10^{2}
1.3.19	2.2±0.4	15±3	1.5±0.1 x10^{5}	4.9±0.6x10^{-4}	160±30	3±1	4.7 x 10^{4}
S223R	2.3±0.1	9±2	2.6±0.8 x10^{5}	n.d.	n.d.	n.d.	>1.8 x 10^{5}
D208G	1.6±0.1	7.3±0.8	2.4±0.7 x10^{5}	0.3±0.1	240±100	2±1 x10^{3}	1.4 x 10^{2}

Table 2.1: Kinetic parameters for the cleavage of substrate **3** & substrate **4** by OmpT, C5 and the four mutants. Reactions were carried out at room temperature.

[1] Substrate sequence: WCARVGKGRGR-NH_2

[2] Substrate sequence: WEEGGRRIGRGGK-NH_2

n.d. No cleavage was observed after a 24 h incubation of 30 μM substrate with 100 nM enzyme.

Enzyme	17	33	87	111	137	149	186	200	208	214	223	288
WT	Ser	Glu	Met	Glu	Ser	Ile	Ser	Tyr	Asp	Asp	Ser	Ile
C5	Ser	Lys	Leu	Val	Asn	Val	Cys	Phe	Asp	Asp	Ser	Phe
1.2.19	Ser	Lys	Leu	Val	Asn	Val	Cys	Tyr	Gly	Val	Ser	Phe
1.3.19	Gly	Lys	Leu	Val	Asn	Val	Cys	Phe	Asp	Asp	Arg	Phe

Table 2.2: List of mutations in WT OmpT, C5, 1.2.19 &1.3.19. The changes in 1.2.19 and 1.3.19 relative to their starting parent, C5, are highlighted in blue.

The 1.2.19 and 1.3.19 proteins were extracted in *n*-octyl-β-glucoside and purified (25). Kinetic analysis of the purified OmpT 1.2.19 and 1.3.19 (Table 2.1) proteins using unlabeled peptide substrates **3** and **4** (Figure 2.2) revealed that the selection yielded the anticipated outcome, namely, highly active enzymes that are specific for the hydrolysis of **3**, in particular cleavage of the Ala-Arg peptide bond, but are impaired in their ability to attack Arg-Arg. It is interesting to note that although cleavage between any two amino-acids in the linker region between the fluorophore and the quencher would lead to increased FL-1 fluorescence, the two best variants 1.2.19 and 1.319 cleaved substrate **3** between Ala-Arg. In addition 1.3.19 did not exhibit the secondary cleavage activities displayed by its parental enzyme C5. Importantly, compared to the wild-type OmpT, 1.3.19 displayed a more than $3x10^6$ reversal in selectivity (Table 2.1) for Ala-Arg over

Arg-Arg cleavage, but its enzymatic activity for cleavage of **3** (k_{cat}/K_M = 1.5x10^5 s^{-1}M^{-1}) was nearly identical to that of the wild-type OmpT for its own preferred substrate **4** (k_{cat}/K_M = 1.7x10^5 s^{-1}M^{-1}). The isolation of both 1.2.19 and 1.3.19 suggests that such highly selective, highly active catalysts, as opposed to enzymes with expanded specificity or low catalytic activity towards new substrates, can be easily isolated by multi-parameter, quantitative screening of large libraries.

Site Specific OmpT Mutants: With the exception of the Ser17Gly mutation in 1.3.19, the other three amino acid substitutions in the two isolated clones resulted in mutations that involved charged residues. Examination of the recently reported structure of OmpT (31) pointed to a significant role for Ser223Arg and Asp208Gly in modulating peptide substrate specificity, since these residues are located deep in the active site cleft.

Ser223Arg and Asp208Gly were constructed using site directed mutagenesis and purified as above. An LC-MS of the substrate cleavage products revealed that the OmpT Ser223Arg and Asp208Gly enzymes cleaved **3** only between Ala-Arg. Neither variant produced the secondary Lys-Gly cleavage that is generated by the wild-type OmpT (k_{cat}/K_M = 4x10^3 s^{-1}M^{-1}) as well as the C5 enzyme. In addition, the single amino acid variants recapitulated the high selectivities and catalytic activities displayed by 1.2.19 and 1.319, respectively (Table 2.1). In particular, Ser223Arg completely abolished the ability of OmpT to cleave Arg-Arg sites (no cleavage of **4** could be detected following >24 hr incubation) yet cleaved **3** at Ala-Arg with a k_{cat}/K_M value slightly higher than

1.3.19. In addition, the Ser223Arg mutant did not cleave substrate **5** (Figure 2.2), further underscoring its specificity for Ala-Arg. The effect of the Ser223Arg mutation appeared to be unique, since insertion of Trp, Lys, Leu, Gly, or Phe at the 223 position using site directed mutagenesis resulted in variants that did not exhibit cleavage of substrate **1** comparable to even wt OmpT when monitored by FACS (Figure 2.4).

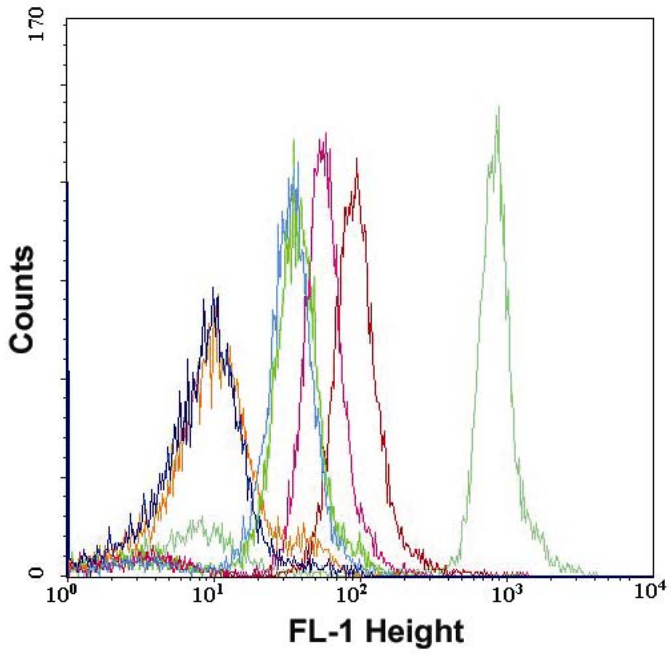


Figure 2.4: Flow-cytometric analysis of the OmpTSer223 mutants. Briefly, the cells were washed and resuspended in 1% sucrose and labeled with 50nM (final concentration) of the AR-FRET substrate **1**. A 20ul aliquot of the labeling reaction was transferred to 0.5ml of 1% sucrose. --- S223W, --- S223L, --- S223F, --- S223K, --- S223G, --- OmpT, --- S223R

2.4 Discussion

We have developed a high throughput strategy for the quantitative screening of enzyme libraries using simultaneous selection and counter-selection criteria. The methodology presented depends on: (1) the display of enzyme libraries on the surface of microorganisms; (2) retention of fluorescent reaction products on the cell surface and (3) multicolor flow cytometry for the isolation of clones that can selectively turn over one or more substrates. The combination of electrostatic and FRET substrates afforded the proper dynamic range required to screen a wide array of enzymatic activities, an important consideration when screening large enzyme libraries.

Display of enzymes on the surface of microorganisms such as *E. coli* or yeast can be accomplished in a variety of ways (32,33). In this manner, the enzyme can react with exogenous synthetic substrates, circumventing the substrate transport limitations associated with methods that use intracellularly expressed enzymes. In our approach, the enzymatic reaction generates fluorescent products that become associated with the cell surface resulting in a fluorescence profile representative of the catalytic selectivity of the displayed enzyme. Although we have capitalized on electrostatic interactions for product capture, a number of other methods for cell surface modification (34,35) may be exploited for the capture of reaction products. In addition to the directed evolution of protease selectivity reported here, we believe that our methodology can be extended to other enzymes including various hydrolases and ligases. For example, studies to engineer

stereo- and enantio- selective variants of the small esterase cutinase are on-going in our laboratory.

We screened an OmpT error-prone library in order to select for variants that preferred a new peptide substrate at the expense of activity with the wild-type preferred substrate. After three rounds of sorting we obtained 10 clones that displayed the desired specificity profile, the best of which were 1.219 and 1.3.19. Although the targeted error-rate for the library was 1.1% (11 bases), both 1.2.19 (Asp208Gly, Asp214Val) and 1.3.19 (Ser17Gly, Ser223Arg) had just two extra mutations each, relative to C5 (20), the starting enzyme construct. A detailed kinetic analysis of the isolated enzymes proved both to be consistent with the flow-cytometric screening criteria. In other words, both of these enzymes have altered as opposed to relaxed substrate specificity, yet maintain a native level of catalytic activity.

Examination of the crystal structure of OmpT (31) indicated that the Asp208Gly mutation in 1.2.19 and the Ser223Arg in 1.3.19 might be the primary determinants of the altered specificity profiles of these enzymes. Construction of these single point mutants of OmpT confirmed that Asp208Gly was similar to 1.2.19 in terms of specificity and overall activity. Unexpectedly, Ser223Arg was more specific than 1.3.19, especially in eliminating cross-reactivity with the wild-type preferred Arg-Arg containing substrate **4**. Importantly, the Ser223Arg variant was as active with substrate **3** as wild-type OmpT with its preferred substrate **4**. Thus, for the Ser223Arg single mutation variant of OmpT, altered specificity has not come at the cost of overall catalytic activity.

It is worth mentioning that while OmpT is a trypsin-like protease (36), the Ser223Arg substitution conferred specificity more analogous to that of chymotrypsin. The effect of this single amino acid substitution should be contrasted with the conversion of trypsin to chymotrypsin that required a monumental effort and hinged on major reorganization of the substrate-binding surface (37,38).

A crystal structure of OmpT has been published (31), but no structural information is available for OmpT containing a bound substrate or substrate analog. Nevertheless, inspection of the OmpT structure leads to a reasonable prediction for the location of an unusually deep S_1 binding pocket (Figure 2.5), with the Ser223 residue located near the bottom of this pocket. Consistent with our proposed role of the 223 residue in substrate recognition, a recent MD calculation using wild-type OmpT and an Arg-Arg containing substrate placed the P_1 Arg side chain deep in the same S_1 pocket proposed here and adjacent to Ser223 (39).

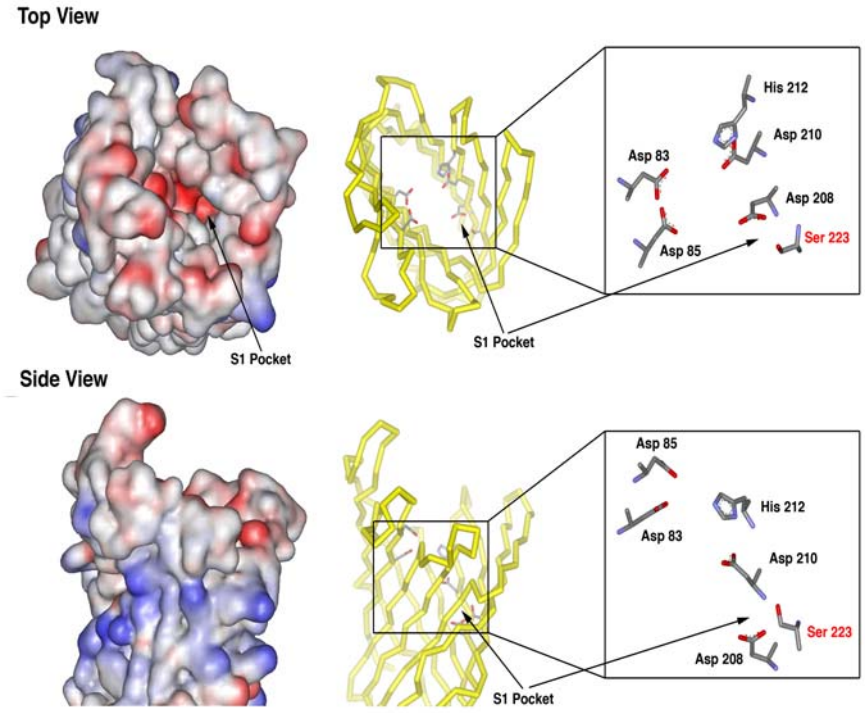


Figure 2.5: (A) The electrostatic potential surface of OmpT (31) generated using WebLab ViewerLite (Accelrys, San Diego, CA) **(B)** The side-chains corresponding to Asp208 and Ser223 are shown in addition to the proposed catalytic residues, Asp83, Asp85, and Asp210 & His212.

The high activity seen with Ser223Arg likely reflects a simple swapping of an important residue from the peptide substrate to the enzyme active site. Because enzymes

evolve to be catalytically active in the presence of substrate, it makes sense that some functional groups attached to the substrate could modulate active site structure and thus catalytic activity to a significant degree. This is the basis of Koshland's induced fit theory of enzyme catalysis (40) in which substrate binding is required to convert an enzyme active site into the proper catalytic arrangement. A logical interpretation of our data is that the Ser223Arg substitution replaces important induced fit interactions of the P_1 Arg side chain of a bound substrate (Figure 2.6) with the 223Arg side chain that is now attached to the enzyme in S_1. A significant structural role for the restored guanidium group would explain why the Ser223Arg variant has wild-type levels of catalytic activity. Interestingly, the decreased rate of catalysis seen for wild-type OmpT reacting with the non-preferred substrate **3** (that lacks the P_1 Arg residue) is primarily a reflection of a substantially lower k_{cat} value, possibly consistent with the induced fit hypothesis that a guanidium group in the P_1-S_1 site is important for high levels of catalytic activity. In order to investigate this hypothesis, we constructed five OmpTSer223 mutants, OmpTSer223Phe, OmpTSer223Gly, OmpTSer223Lys, OmpTSer223Leu and OmpTSer223Trp. These five mutants were designed to test if the guanidium group is essential in P_1-S_1 site for high catalytic activity or if steric (Phe, Trp), hydrophobic (Leu) or positively charged (Lys) residues can help achieve the same goal. Flow-cytometric analysis (Figure 2.4) demonstrates that none of these mutants had even a wild-type level of activity with substrate **1**, confirming the need for a guanidium group. On-going studies will elucidate whether the generation of equally active enzymes selective for different S_1

occupancies, for example cleavage of Asp-Arg, Trp-Arg, or Arg-Ala sequences, can be accomplished by similar compensatory changes in the OmpT active site.

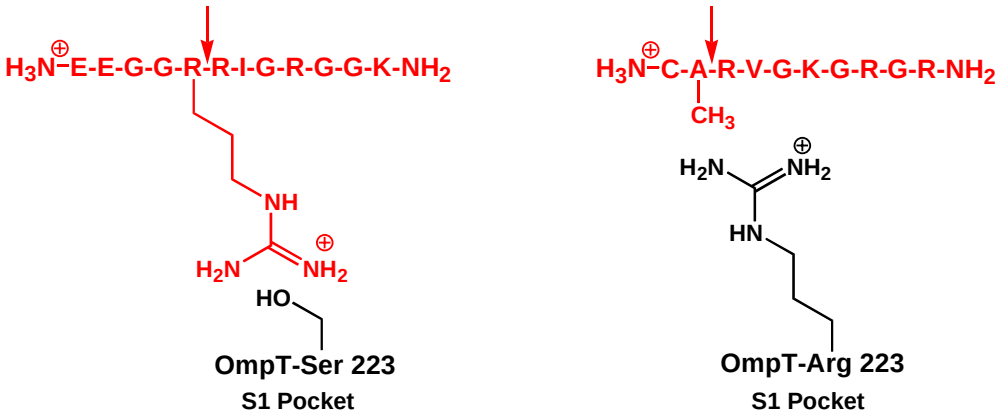


Figure 2.6. Arginine side-chain swapping. The Arg occupying the P1 subsite is accommodated by Ser223 in S1. In the 1.3.19 mutant the Ser223Arg only allows occupancy by Ala in the P1 site of the substrate.

Swapping important residues at protein-protein and protease-substrate interfaces has been accomplished by rational design (41,42). For 1.3.19, this apparent swapping occurred as a result of combined selection and counter-selection. In that regard, the isolation of the OmpT Ser223Arg variant supports the notion that residue swapping at protein-peptide interfaces represents a facile mechanism for substrate diversification as

the result of evolutionary pressure without any preconceived notions regarding the protein structure.

Impressive as our findings with OmpT Ser223Arg might be, in general, the engineering of large changes in substrate specificity is expected to require multiple amino acid substitutions leading to extensive remodeling of the active site. Thus, while in this case, the Ser223Arg mutant could have been identified by screening a small library using manual techniques, the evolution of mutants of OmpT or other enzymes that can react selectively with non-native substrates will likely require the screening of very large libraries. Therefore, a high overall assay throughput/dynamic range and the ability to place selection and counter-selection evolutionary pressure on an enzyme to be evolved are key to the isolation of rare, change-of function clones. Selection and counter-selection assays can be carried out in a tiered approach, wherein the library is screened first for one substrate and then the active clones are screened for activity with the counter-selection substrate. As shown here, however, selection and counter-selection can also be applied simultaneously, analogous to living systems. Simultaneous screening by FACS accommodates high throughput and is particularly attractive for carrying out evolutionary experiments using three or more substrates, as required to explore "substrate space".

2.5 Conclusions

We have a developed a new two-color flow-cytometric dual-color selection and counter-selection assay for the quantitative screening of libraries the *E. coli* surface displayed protease OmpT. The application of this assay towards altering the P1 specificity of OmpT to hydrolyze Ala↓Arg substrate was demonstrated. Although the engineered enzyme had just one amino acid change, it showed a 3 million-fold change in selectivity (-Ala-Arg-/-Arg-Arg-) and a catalytic efficiency for Ala-Arg cleavage that is the same as that displayed by the parent for the preferred substrate, Arg-Arg. The engineered variant was thus highly active and specific.

2.6 References

1. Carter, P., Wells, J.A. (1987) *Science* **237**, 349-.

2. Bone, R., Fujishige, A., Kettner, C.A. , Agard, D.D. (1991) *Biochemistry* **30**, 10388-10398.

3. Palzkill, T., Botstein, D. (1992) *J Bacteriol.* **174**, 5237-5243.

4. Graham, L.A., Brandt, U., Sargent, J.S., Trumpower, B.L. (1993) *J Bioenerg. Biomembr*. **25**, 245-257.

5. Harris, J.L., Craik, C.S. (1998) *Curr Opin Chem Biol*. **2**, 127-132.

6. Farinas, E.T., Butler, T., Arnold, F.H. (2001) *Curr Opin Biotechnol*. **12**, 545-551.

7. Arnold, F.H. (2001) *Nature,* **409***,* 253-7.

8. Hedstrom, L. (2002) *Chem Rev*. **102**, 4501-4524.

9. Tao, H., Cornish, V.W. (2002) *Curr Opin Chem Biol*. **6**, 858-864.

10. Matsumura, I., Ellington, A.D. (2001) *J Mol Biol*. **305**, 331-339.

11. Geddie, M.L., Matsumura, I. (2004) *J Biol Chem*. **279**, 26462-26468.

12. Aharoni, A. Gaidukov, L., Khersonsky, O., McQ Gould, S., Roodveldt C., Tawfik, D.S. *Nature Genetics* **37**:73-6 (2005).

13. Reetz, M.T., Wilensek, S., Zha, D. and Jaeger, K-E. (2001) *Angew Chem Int Ed* **40**, 3589-3591.

14. DeSantis, G., et al., (2003) *J Am Chem Soc,* **125**, 11476-11477.

15. Van Loo, B, Spelberg JH, Kingma J, Sonke T, Wubbolts MG, Janssen DB. (2004) *Chem Biol* **11**, 981-990.

16. Rothman, S.C., Kirsch, J.F. (2003) *J Mol Biol*. **327**, 593.

17. Yano T and Kagamiyama, *Proc. Natl. Acad. Sci. U.S.A* **98**,903-907 (2001)

18. Santoro, S.W., Schultz, P.G. (2002) *Proc Natl Acad Sci USA*. **99**, 4185-4190.

19. Bucholz, F., Stewart, A.F., *Nat Biotechnol*. **19**, 1047-52 (2001).

20. Olsen, M.J., Stephens, D., Griffiths, D., Daugherty, P., Georgiou, G., Iverson, B.L. (2000) *Nat Biotechnol*. **18**, 1071-1074.

21. Francisco, J.A., Campbell, R., Iverson, B.L., Georgiou, G. (1993) *Proc Natl Acad Sci USA*. **90**, 10444-10448.

22. Georgiou, G., Stathopoulos, C., Daugherty, P.S., Nayak, A.R., Iverson, B.L., Curtiss, R. 3rd (1997) *Nat Biotechnol.* **15**, 29-34.

23. Kukkonen, M., Korhonen, T.K. (2004) *Int J Med Microbiol*. **294**, 7-14.

24. Okuno, K., Yabuta, M., Ooi, T., Kinoshita, S. (2004) *Appl Environ Microbiol*. **70**, 76-86.

25. McCarter, J.D., Stephens, D., Shoemaker, K., Rosenberg, S., Kirsch, J.F., Georgiou, G. (2004) *J Bacteriol*. **186**, 5919-5925.

26. Dekker, N., Cox, R.C., Kramer, R.A., Egmond, M.R. (2001) *Biochemistry* **40**, 1694-1701.

27. Okuno, K., Yabuta, M., Kawanishi, K., Ohsuye, K., Ooi, T., Kinoshita, S. (2002) *Biosci Biotechnol Biochem*. **66**, 127-134.

28. Matthews, D.J., Wells, J.A. (1993) *Science* **260**, 1113-1117.

29. Razatos, A., Ong, Y.L., Sharma, M.M., Georgiou, G. (1998) *Proc Natl Acad Sci USA*. **95**, 11059-11064.

30. P.S. Daugherty, B.L. Iverson, G. Georgiou (2000) *J Immunol Methods* **243**, 211-227.

31. Vandeputte-Rutten, L., Kramer, R.A., Kroon, J., Dekker, N., Egmond, M.R., Gros, P. (2001) *EMBO J*. **20**, 5033-5039.

32. Lee, S.Y., Choi, J.H., Xu, Z. (2003) *Trends Biotechnol*. **21**, 45-52.

33. Chen, W., Georgiou, G. (2002) *Biotechnol Bioeng*. **79**, 496-503.

34. Saxon, E., Bertozzi, C.R. (2001) *Annu Rev Cell Dev Biol*. **17**, 1-23.

35. Link, A.J., Vink, M.K., Tirrell, D.A. (2004) *J Am Chem Soc*. **126**, 10598-10602.

36. Leytus, S.P., Bowles, L.K., Konisky, J., Mangel, W.F. (1981) *Proc Natl Acad Sci USA*. **78**, 1485-1489.

37. Hedstrom, L., Szilagyi, L., Rutter, W.J. (1992) *Science* **255**, 1249-1253.

38. Kurth, T., Ullmann, D., Jakubke, H.D., Hedstrom L. (1997) *Biochemistry* **36**, 10098-10104.

39. Baaden, M., Sansom, M.S. *Biophys J.* , in press (available at http://www.biophysj.org/cgi/rapidpdf/biophysj.104.046987v1.pdf).

40. Koshland, D.E., (1958) *Proc Natl Acad Sci USA*. **44**, 98-104.

41. Atwell, S.,Ulstch, M.,De Vos, A.M., Wells, J.A. (1997) *Science* **278**, 1125-1128.

42. Caputo, A., Parrish, J.C., James, M.N., Powers, J.C., Bleackley, R.C. (1999) *Proteins* **35**, 415-424.

Note : BODIPY® is a registered trademark of Molecular Probes (Eugene, OR).

Chapter 3

High-throughput protease engineering

3.1 Introduction

Proteases are ubiquitous proteins that play a key role in a wide range of physiological processes such as protein turnover(1), apoptosis (2) and signaling, (3). Proteases represent one of the most well-studied and characterized enzyme classes (4-6). In spite of the great body of work dedicated to the understanding of the mechanistic aspects of protease catalysis (7) and their substrate specificity (8), designing high-throughput assays for the engineering of protease specificity has been challenging. The two main obstacles for high-throughput protease screening *in vivo* are expression and selection. Expression of soluble proteases *in vivo* is hampered by cytotoxicity, due to cleavage of essential physiological proteins (9) and auto-proteolysis. And unlike the selection for affinity, wherein binding maintains the link between genotype and phenotype (For a review on *in vitro* protein evolution please see (10)), designing high-throughput selection for catalytic turnover in a quantitative manner is significantly more complex.

OmpT is a non-essential *E. coli* outer membrane endopeptidase that cleaves between dibasic residues (11) and has been implicated in pathogenicity (12). OmpT is a trypsin-like protease in that the requirement for Arg in P1 is strict, while it can tolerate multiple

residues in P1' (although Arg is still preferred) (11). Similar to other *E. coli* outer membrane proteins, the protein is translated with a leader peptide, translocated across the cytoplasmic membrane and folded into its final active conformation in the outer-membrane (13). The catalytic core of OmpT is directed towards the external environment and thus has no access to the soluble proteins of the cell (14).

We had previously reported a high-throughput flow-cytometric, dual color, selection and counter-selection assay for the directed evolution of OmpT variants (Figure 2.1, Chapter 2) (15). The selectivity profile of the outer-membrane protease gives rise to a multi-color profile. Flow-cytometry is then used to isolate cells that show the desired fluorescence characteristics. In the current work, the application of the two-color flow-cytometric assay to systematically engineer the substrate specificity of OmpT is demonstrated.

3.2 Materials and Methods

Substrate synthesis.

The peptides **1a-5a** Ac-CXRVGKGRGR-NH_2 (X=A/P/T/E/Y), **7a** Ac-CEAVGKGRGR-NH_2 and **6a** Ac-EEGGRRVGKGRGR-NH_2, were synthesized at the University of Texas Peptide synthesis facility. QSY®7 C_5 maleimide, 5-carboxytetramethylrhodamine, succinimidyl ester (5-TAMRA, SE) and BODIPY®-FL-SE were purchased from Invitrogen (Carlsbad, CA). The peptides **1-5** WXARVGKGRGR-NH_2 (X=A/P/T/E/Y), **7** WCEAVGKGRGR and **6**

WEEGGRRVGKGR-NH_2 (>95% purity each) were purchased from EZBiolab (Westfield, IN).

The conjugation of fluorophores to the substrates **1a-7a** was performed essentially as described before (15).

pDMLE19 plasmid construction: In order to construct an expression vector for the cloning of OmpT libraries downstream of its native promoter**,** the EcoRI restriction site in pUC19 was removed as follows. 5ug of pUC19 was digested with 25U of EcoRI in a 50µl reaction incubated at 37^{o}C for 3h. The product was gel-purified incubated with 33uM (final concentration) dNTPs and 1.25U of the Klenow fragment of the DNA polymerase in a 50ul reaction volume at 25^{o}C for 15min. At the end of the reaction, 10µl of 50mM EDTA was added and the enzyme heat inactivated by incubation at 75^{o}C for 20min. The product was gel-purified and blunt-end self-ligated using T4 DNA ligase, transformed and sequenced to confirm removal of the EcoRI site and labeled pDUCE19. The ompT upstream region containing its native promoter was amplified using the primers 5' CGCGGATCCGATTCGAACCTGCGGGC 3' and 5' CCGGAATTCTTAAAATGTGTACTTAAC 3'. The ompT gene was amplified from plasmid pML19 (16)(REF) as the template with the primers 5' AAACGATTGAATGGAGAATTCACCATCGGGCGAAACTTCTG 3' & 5' AACAGCCAAGCTTTTAAAATGTGTACTTAAGACCAGCAGT 3'.The two different PCR products were gel-purified and 50ng of each fragment was mixed together and

reamplified using the primers 5' CGCGGATCCGATTCGAACCTGCGGGC 3' and 5' AACAGCCAAGCTTTTAAAATGTGTACTTAAGACCAGCAGT 3'.The resulting PCR product was digested with BamHI & HindIII and ligated into pDUCE19 cut with the same restriction enzymes. Transformation and sequencing confirmed the construct pDMLE19. This construct has the EcoRI site right before the ATG start site allowing easy cloning of the libraries.

Library construction

Error-prone libaries: Random mutagenesis was performed by error-prone PCR amplification using the ompT gene/ompT mutants as the template and the primers 5' CCGGGAATTCACCATGCGGGCGAAACTTCTGGGAATAGTC 3' and 5' AACAGCCAAGCTTTTAAAATGTGTACTTAAGACCAGCAGT 3'. The error-rates were varied between 0.6%-1.0%. Denaturation was performed by heating to 95^{o}C, 2 min 16 cycles – 95^{o}C, 1 min / primer annealing was performed at 52^{o}C, 1 min / and primer extension was carried out at 72^{o}C, 3 min; 72^{o}C, 10 min. The PCR products were run on a 0.8% agarose gel and purified using the QIAGEN QIAquick gel extraction kit. The purified DNA was digested with EcoRI and HindIII overnight at 37^{o}C and heat inactivated at 65^{o}C for 20 min. Digested product was run on a 0.8% agarose gel, QIAGEN purified and the amount of DNA quantified. The vector pDMLE19 was similarly digested, dephosphorylated and purified. Ligation reactions were performedusing 100 ng (50 fmol) of the digested vector and a 3-fold molar excess of the

insert at 25°C for 4 h. The ligation reactions were heat inactivated at 65°C for 20 min and desalted on a nitrocellulose membrane for 1h. A 3 µl aliquot of the desalted ligation reaction was transformed with 40 µl of electrocompetent MC1061 (*araD139* Δ (*ara-leu*) 7696 Δ*lac174 galU galK hsr*$^-$ *hsm*$^+$ *strA*^{R}) cells. Following incubation in 1ml SOC at 37°C for 1 h the entire transformation mixture was plated on LB 1% glucose plates containing 200 µg/ml ampicillin to yield ~2x10^8 independent transformants. After growth at 37°C for 8 h the cells on plates were scraped using 2 ml LB, pooled together and 1 ml of the pooled, scrapped cells was used to inoculate 1 L of LB 1% glucose containing 200 µg/ml ampicillin. The cells were grown in LB/ampicllin 1% glucose at 25°C for 5 h. Plasmid DNA was then isolated, re-transformed into electrocompetent *E. coli* BL21(DE3) (F$^-$ *ompT hsdS*$_B$(r_B^- m_B^-) *gal dcm* (DE3)) and the cells grown to an OD_{600} of 2 units in 1 L of 2xYT containing 200 µg/ml of ampicillin at 37°C for library screening.

NNS libraries: The entire ompT gene was assembled based on a set of 48 primers designed to include a 5' EcoRI cut-site and a 3' HindIII cut site for ligation into the pDMLE19 vector. Denaturation of the double-stranded template was performed at 95°C, 1 min; 54 cycles – 95°C, 30s /primer annealing was carried out at 45°C, 30s / and primer extension was performed at 72°C, 30 s. A 2 µl aliquot of the first PCR reaction was used as template to for a second PCR reaction, amplified using the outside primers 5' CCGGGAATTCACCATGCGGGCGAAACTTCTGGGAATAGTC 3' and 5' AACAGCCAAGCTTTTAAAATGTGTACTTAAGACCAGCAGT 3'. Denaturation of the double-stranded template was performed at 95°C, 2 min; 29 cycles – 95°C, 30 s /

primer annealing was carried out at 60^{o}C, 30 s / and primer extension was performed at 72^{o}C, 90s; 72^{o}C, 10 min. The PCR product was then gel-purified, digested and ligated into pDMLE19 as described above. In order to construct the saturation library, Sat3, PAGE-purified primers containing NNS at codons 27,208 and 223 were used for the gene assembly PCR reaction. The library Sat7 was constructed using randomized NNS primers encoding for residues 27, 29, 39, 40, 42, 208 and 223 in the assembly PCR reaction. For the construction of the 90% NNS library, a 9:1 mix of the NNS primers and the wild-type codon primers corresponding to residues 27, 29, 39, 40, 42, 44, 81, 87, 97, 101, 148, 150, 153, 163, 170, 208, 221, 263, 265, 280 and 282 were used to assemble the gene.

Library Screening: The plasmid library (construction and isolation of plasmid library described above) was transformed into electrocompetent *E. coli* BL21(DE3) (F^- *ompT* *hsdS*$_B$(r_B^- m_B^-) *gal dcm* (DE3)) and the cells grown to an OD_{600} of 2 units in 1 L of 2xYT containing 200 µg/ml of ampicillin at 37^{o}C (6-8 h). An aliquot of the cells adjusted to an OD_{600} of 2 units (1 ml) were centrifuged at 10Krpm for 2 min using a micro-centrifuge. The supernatant was discarded and the cells were resuspended in 1 ml of 1% sucrose. Following resuspension, the cells were again spun down at 10 Krpm and resuspended in 1ml 1% sucrose. A 5 0µL aliquot of the cell suspension was added to 948 µl of 1% sucrose and labeled using 1 µl of the appropriate selection substrate (final concentration 20 nM) and 1µl of the counter-selection substrate Arg-Arg (final concentration 100 nM). An 800ul aliquot of this labeling reaction was diluted into 200 µl of 1% sucrose and

analyzed on the flow-cytometer. Library sorting was performed on a Dako MoFlo® instrument, using gates set based upon FSC/SSC and FL1/Compensated FL2 values. A total of ~$2x10^7$ cells were screened using three separate labeling reactions and ~1% of the most fluorescent cells were sorted into 3 ml of 2xYT. The cells were grown to an OD_{600} of 2 units, resorted and the process repeated until the mean fluorescence of the highly fluorescent population remained constant. After 5-7 rounds of sorting and resorting the cells were plated on selective media and individual colonies were analyzed on the flow-cytometer.

Flow-cytometric OmpT activity assays: Single colonies from a plate were used to inoculate 2 ml 2xYT cultures containing 200 μg/ml of ampicillin. The cells were grown to an OD_{600} of 2 units, centrifuged at 10 Krpm for 2 min, washed once with 1 ml of 1% sucrose and resuspended in 1ml of 1% sucrose. For labeling, 50 μl of the cells were diluted into 949 μl of sucrose and 1μl of the substrate (final concentration 20 nM for **1-5** and 100nM for **6**). A 20 μl aliquot of the labeling reaction was diluted into 0.5 ml of 1% sucrose and analyzed on Becton Dickinson FACSSort.

Enzyme purification and kinetic analysis. Proteins were isolated as previously described, with minor modifications (15) to a final purity >90% as determined by SDS-PAGE. For kinetic analyses, 0.5-20 nM of the purified enzymes were incubated with 20 μM to 1 mM of the appropriate substrate in 0.1 M Tris, 50 μM TCEP, 10 mM EDTA, pH

6.1 at room temperature (25°C) and the reaction was monitored by HPLC on a Phenomenex C18 reverse phase column using the following gradient: 5% AcN/95% H_2O for 1 min, increasing to 95% AcN/5% H_2O over a period of 29 mins and returning to 5% AcN/95% H_2O over 5 mins. The product concentration was determined using the integration areas at 280 nm and the apparent rates were fitted to a Michaelis-Menten by non-linear regression. The cleavage products were determined by LC-MS (ESI) on a Magic 2002 instrument (Micron Bioresources, Auburn, CA).

Plasminogen activation assays: Stock solutions of Spectrozyme ®PL (5.0 mM) and human glu-type plasminogen (0.5 mg/ml) were made in 50 mM Tris, 0.01% Tween 20, pH 7.4. To 44 ml of 50 mM Tris, 0.01% Tween 20, pH 7.4 was added 5ml of Spectrozyme PL (final concentration 0.5 mM) and 1 ml of plasminogen to make up the working solution. To 200 µl of the working solution was added 0.2-5.0 µg of the purified enzyme (vtPA (American Diagnostica, Stamford, CA)) /eRV6/OmpT) and the reaction kinetics monitored at 405nm on a BioTek Synergy HT (BioTek instruments, Winooski, VT) at 37°C. In order to estimate if the enzymes had any activity towards hydrolyzing the chromogenic substrate, identical assays were run in the absence of plasminogen.

T7 RNA polymerase digests: 300 U of T7 RNA polymerase (Ambion Inc, Austin, TX) was incubated with ~50-200 ng of OmpT/OmpT variants 10 mM MES (2-morpholinoethanesulfonic acid), 10 mM EDTA, pH=6.1 in a 30µl reaction volume at

37^{o}C for 3 h. 15µl of the reaction mixture was mixed 15 µl of 2X SDS page loading buffer containing DTT and run on a 12% Tris-glycine gel. The sites of cleavage were confirmed using N-terminal sequencing.

Site-directed mutagenesis: Site-directed mutagenesis was performed using the Quikchange ® (Stratagene, La Jolla, CA) strategy using pDE19 as the template and the following primer pairs to introduce the appropriate mutations:
E27Lf 5' GACATTAGTCTTGGAACTCTGAGCGGAAAAACAAAACTGC 3' & E27Lr 5' CCTCCTTCTTCGGCTAGATAAACACGCAGTTTTGTTTTTC 3'; D208Rf 5' CTGGGTGGAATCATCTCGTAACGATGAACACTATGACCCG 3' & D208Rr 5' GTTACGAGATGATTCCACCCAGCCGCTGTATTTAAATGTG 3'; S223Gf 5' GGAAAAAGAATCACTTATCGCGGTAAGGTCAAAGACCAAAAT 3' & S223Gr 5' CATTGACTGCAACAGAATAGTAATTTTGGTCTTTGACCTTACCG 3'.

Determination of the relative substrate specificities: 30 µM of the appropriate substrate was incubated with 1-20 nM enzyme for 45min such that the enzyme cleaved the preferred substrate 40-50%. Reaction conditions were optimized for each enzyme-substrate pair and the percentage cleavage was estimated on the HPLC. The relative substrate preferences are then reported as a percentage/ratio of the preferred substrate cleavage.

Human-β-defensin 3 assays: *E. coli* viability assays were performed using the liquid microdilution assays in salt-free buffers (17). Briefly, cells were grown to an OD_{600} of 2 units in 2xYT; 0.5 ml cells spun down at 10 krpm for 2 min, washed and resuspended in 0.5 ml 10m M Na_2HPO_4, pH = 7.4. 9µl of cells (~10^7 cells) were mixed with 1 µl of human-β-defenesin3 (final concentration 40ug/ml) or 1 µl water (control) and incubated at 37°C for 4 h. The cells were then recovered by the addition of 990 µl of LB and serial dilutions plated to estimate cell viability. The reported percentages are the number of cells surviving in the presence of the defensin to the corresponding ratio of cells incubated with water as a control. Human-β-defensin3 (~90% pure) was purchased from AnaSpec (San Jose, CA).

3.3 Results

3.3.1 Engineering the P1 specificity of OmpT

Modification of the selection substrate: The FRET selection substrate described in Chapter 2 for the two-color selection/counter-selection flow-cytometric assay used tetramethylrhodamine (TMR, red) to quench BODIPY (BD, green) fluorescence. Substrate cleavage results in release of the TMR quencher, and the captured cleavage product labels the cell with green fluorescence. Since the intact FRET selection substrate has an overall positive charge, it is anchored on the negatively charged *E. coli* surface and cells not expressing OmpT should be decorated with the uncleaved FRET substrate. However, since TMR quenches the BD fluorescence, and the quenching causes excitation

and subsequent fluorescence of TMR, cells not expressing OmpT have a red (TMR) fluorescent profile. This is not a problem with cells that do not express OmpT or cells that express inactive OmpT variants, as these will show up as red fluorescent cells and will not be selected. Selection of moderate activity (cells expressing enzyme that cleave only a fraction of the FRET substrate anchored on the surface) but specific OmpT variants expressed on the surface cannot be achieved using the current assay. Since the enzyme is able to turnover only part of the FRET selection substrate, these cells will be decorated with a combination of the fluorescent BD fragment (green fluorescence) and the uncleaved substrate (red fluorescence). The fluorescence profile of these cells is thus both green and red, the same as non-specific enzymes. The ability to monitor fluorescence and hence reactivity with the WT counter-selection substrate labeled with TMR is thus lost (Figure 3.1).

In order to overcome this limitation, we explored the use of a non-fluorescent quencher. QSY7 (Invitrogen, Carlsbad, CA), a dark quencher, has spectral properties ideally suited for overlap with BD. Conjugation of QSY7 –C_5-maleimide to the cystiene on the Ala↓Arg selection peptide under basic conditions (0.5 M $NaHCO_3$) yielded the conjugate, AR-Q7. Following FPLC purification, AR-Q7 was conjugated to BODIPY-FL-SE (Invitrogen, Carlsbad, CA) using the lysine ε-amine on the peptide under mildly basic conditions (10mM Na_2HPO_4, pH = 7.4) to yield AR-BQ7. The product was FPLC

purified using a reverse-phase column and product identity confirmed using mass-spectrometry.

Cells expressing no enzyme, *E. coli* BL21(DE3) [*ompP ompT*]; WT OmpT; and OmpTS223R (a highly active Ala↓Arg enzyme, Chapter 2) were grown up in 2ml cultures with ampicillin to an OD_{600} of 2 units at 37^{o}C. 1 ml of each culture was washed with 1% sucrose to get rid of the salts and resuspended in 1 ml 1% sucrose. An aliquot of the cells were then labeled with either 20 nM AR-FR (Fluorophore = BD/ Quencher = TMR) or 20 nM AR-BQ7 (Fluorophore = BD/ Quencher = QSY7) and analyzed by flow-cytometry. All three populations showed similar profiles with both probes (not shown), indicating no significant differences between TMR and QSY7 as the quenching partner in detecting highly active enzymes.

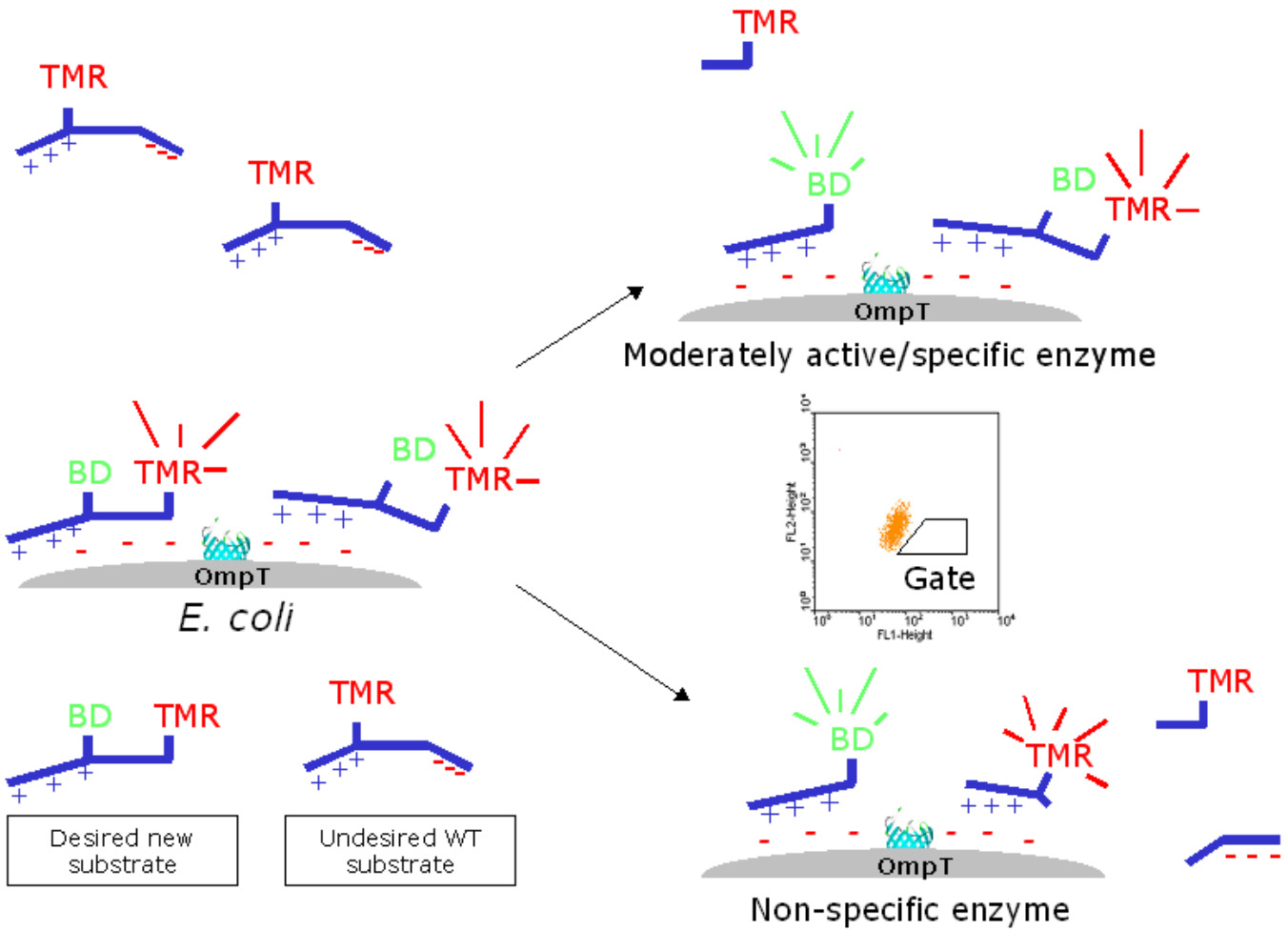


Figure 3.1: Two-color flow-cytometric scheme that illustrates the inability of the BD/TMR FRET selection substrates to distinguish between moderately active but selective enzymes and non-specific enzymes.

In order to evolve a set of highly active OmpT endopeptidases that have P1 specificities altered systematically to recognize one amino acid from each of the six classes of amino acids (Table 3.1) (18), a set of random mutagenesis and multiple residue saturation libraries were constructed using error-prone PCR and primer based assembly mutagenesis (Table 3.2).

Amino acid	Property	Targets for OmpT evolution	
		P1	P1'
Arginine	Basic	Arg-Arg (WT)	Arg-Val
Alanine	Hydrophobic	Ala-Arg	
Glutamic Acid	Acidic	Glu-Arg	Glu-Ala
Tyrosine	Aromatic	Tyr-Arg	
Threonine	Polar	Thr-Arg	
Proline	Folding	Pro-Arg	

`

Table 3.1: Targets for the directed evolution of OmpT

Library	Size	Mutagenesis
1ePCR	2 x10^8	Error-prone : 1.0%
3ePCR	1 x10^8	Error-prone : 3.0%
50% NNS	2 x10^8	NNS : 21 residues, 50% randomized primer *
90% NNS	3 x10^8	NNS : 21 residues, 90% randomized primer *
Sat3	1 x10^6	Saturation : 27,208, 223
Sat7	6 x10^8	NNS : 27,29,39,40,42,208 & 223
Sat3eP	2 x10^8	Error-prone 1% on Sat3 plasmid

Library	Size	Mutagenesis	Parent
ER3ePCR	1 x10^8	Error-prone 1%	StER3
ER3Sat3	2 x10^6	Saturation: 89,97,150	StER3
TRePCR	2 x10^8	Error-prone 0.8%	StTR2
PR5ePCR	3 x10^8	Error-prone 0.8%	90PR3
S223DePCR	2 x10^8	Error-prone 0.6%	S223D

Table 3.2: The set of libraries used for the engineering of OmpT substrate specificity. The assembly reaction was carried out using a primer mixture with either a 1:1 or 9:1 ratio of randomized primer to the wild-type (WT) primer. For more details on the PCR assembly reactions please refer to Materials and Methods.

Engineering of Glu↓Arg activity: In Chapter 2 we reported the evolution of an OmpT variant that efficiently cleaved an Ala-Arg substrate but not the WT Arg-Arg

subtrate(15). We then attempted to isolate an enzyme variant that would selectively recognize a Glu↓Arg substrate while displaying no WT (Arg-Arg) activity. An error-prone library, 1ePCR (Table 3.2), was constructed by mutagenic PCR based on the *ompT* gene as template. The entire library was cloned into pDUCE19 (pUC19 based vector designed to express OmpT under the control of its native promoter). The ligation mixture was transformed into electrocompetent *E. coli* MC1061 to yield $2x10^8$ independent transformants. Plasmid was pooled and retransformed into the selection strain, *E. coli* BL21(DE3), which is an *ompT ompP* deficient strain. The cells were grown for 6-8 h at 37°C until they reached an OD_{600} of 2 units. A 1 ml aliquot of the culture (~10^9 cells) was washed and resuspended in 1% sucrose to get rid of salts and labeled with 20 nM Glu↓Arg and 100 nM Arg↓Arg, then sorted on the MoFlo (Dako, Fort Collins, CO) (Figure 3.2) flow-cytometer. Unfortunately, after five rounds of sorting, no enrichment was observed.

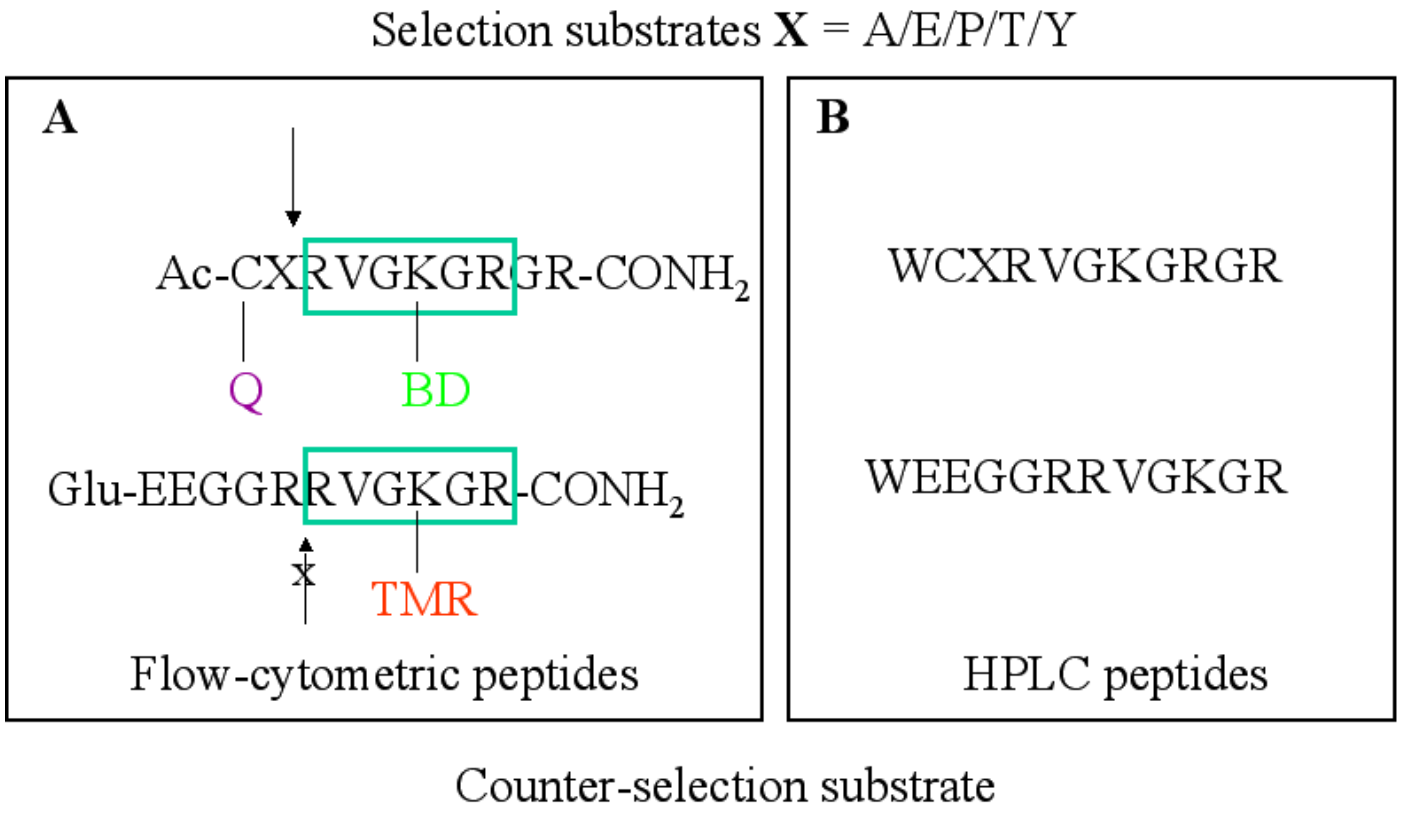


Figure 3.2: (A) Peptide sequence of the selection and counter-selection substrates. The amino acid sequence in green is common to both the selection and counter-selection substrates and hence enzymes that cleave the selection substrate at amino acids other than the targeted Xaa↓Arg site will also cleave the counter-selection substrate and show up as non-specific enzymes on the flow-cytometric assay. The current selection scheme thus allows for isolation of enzymes variants that selectively cleave at the targeted Xaa↓Arg site while simultaneously deselecting against Arg↓Arg. Q = QSY7 (quencher); BD = BODIPY; TMR = Tetramethylrhodamine (B) Amino acid sequence of the selection and counter-selection substrates used in the HPLC assay (Trp absorbance at 280nm) for kinetic characterization of the enzymes

The lack of success in isolating variants with altered Glu↓Arg specificity was not surprising given the fact that electrostatic reversal of ion-pair specificity had been attempted in multiple proteins including trypsin with very little or no success (19, 20). In fact the ion-pair reversal by protein engineering had been deemed unlikely in the literature because of completely different dielectric constants for the two different microenvironments (21). To circumvent this problem, a complete remodeling of the S1 subsite was required to change the microenvironment around the charge. Accordingly, a saturation mutagenesis library of the three residues (Glu27, Asp208, Ser223) that constitute the bottom of the S1 binding site of OmpT was constructed using assembly PCR. The PCR product was similarly cloned and transformed into electrocompetent *E. coli* BL21(DE3) to yield $1x10^6$ independent transformants. The *E. coli* OmpT library was washed and resuspended in 1% sucrose and labeled with 20 nM Glu↓Arg selection substrate and 100 nM Arg↓Arg counter-selection substrate (Figure 3.2). Cells expressing high BODPIY fluorescence (green, FL-1, 530/40 emission) and low TMR fluorescence (red, FL-2, 570/40 emission) were sorted into 3 ml of 2xYT on the MoFlo flow-cytometer. The cells were grown to an OD_{600} of 2 units at 37^{o}C, labeled and sorted again. After four rounds of sorting (Figure 3.3), no further increase in the mean fluorescence was observed, and the sorted cells were plated onto selective agar plates. Single colonies were picked and their fluorescence profile analyzed separately using the selection and counter-selection substrates.

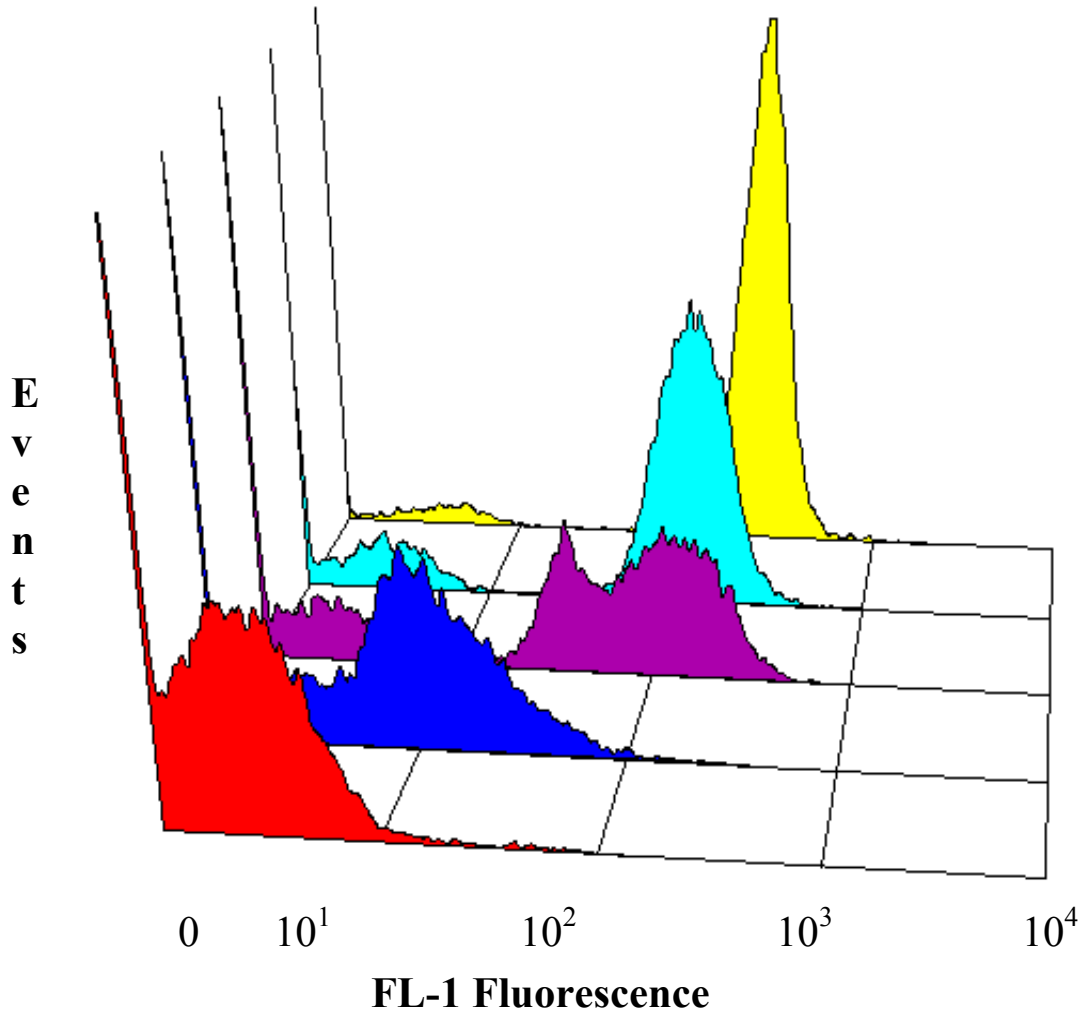
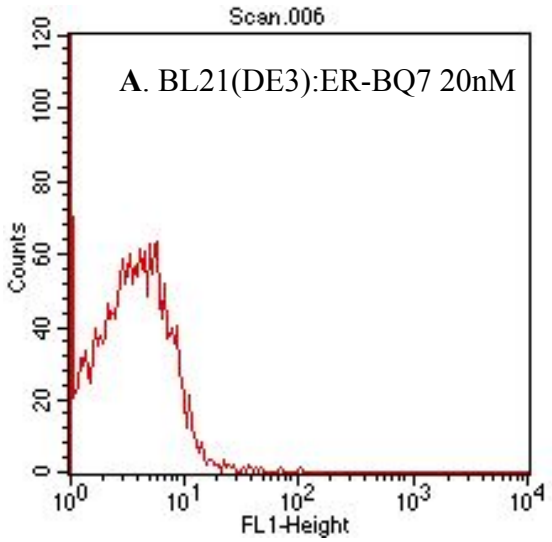
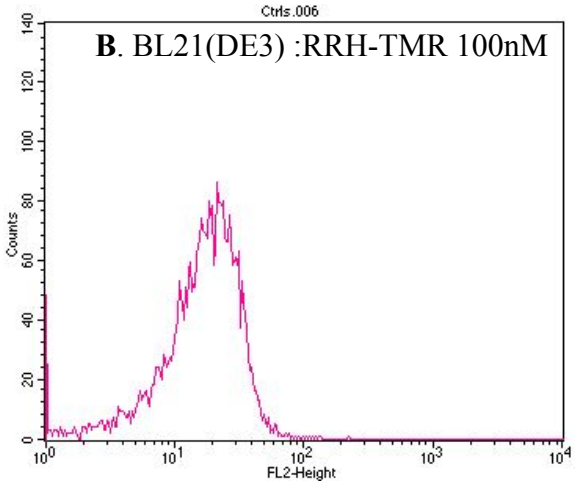


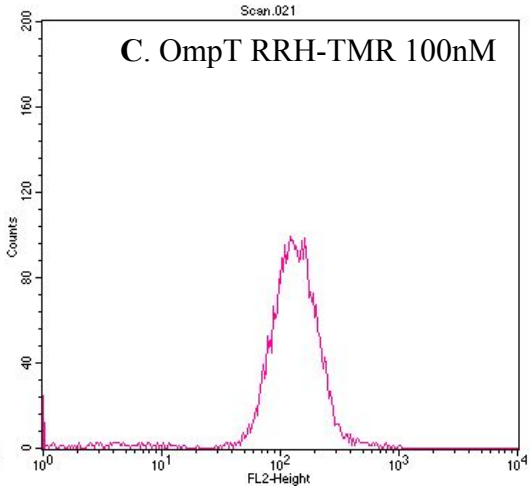
Figure 3.3: Single-color fluorescence profile of the *E. coli* OmpT Sat3 library though multiple rounds of sorting using the Glu↓Arg substrate.

Five clones that showed high FL-1 fluorescence and low FL-2 fluorescence (Figure 3.4), consistent with the sorting criteria, were sequenced. DNA sequencing of the gene that

encodes the high-fluorescence variants indicated that all five clones were the same OmpT variant (designated StER3).







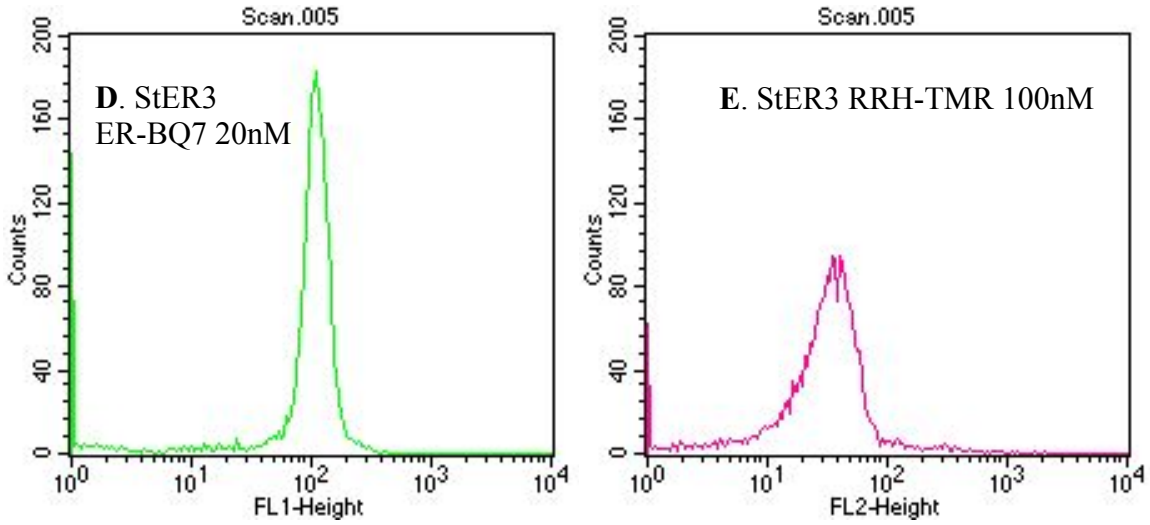


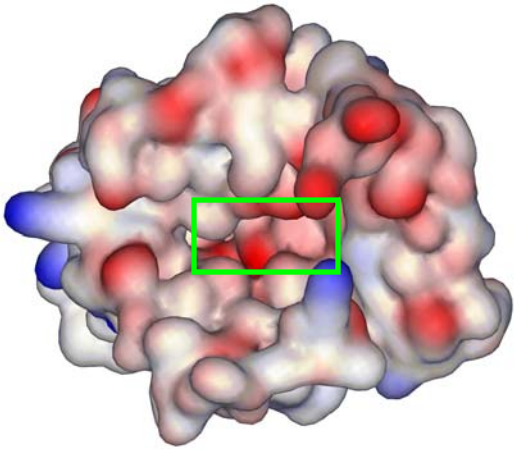
Figure 3.4: Fluorescence histograms of (**A&B**) BL21(DE3), (**C**) wild type OmpT and (**D&E**) StER3 with the Glu↓Arg (selection) and Arg↓Arg (counter-selection) substrates. Briefly, cells corresponding to an OD of 2 were spun down, washed and resuspended in 1% sucrose. 50 µl cells diluted to 1 ml in 1% sucrose were labeled with the indicated substrate and an aliquot of the labeling reaction analyzed on the flow-cytometer.

Enzyme	Residue number		
	27	208	223
OmpT	Glu	Asp	Ser
StER3	Leu	Arg	Gly

Table 3.3: List of amino acid changes in the Glu↓Arg variant, StER3, relative to OmpT. Acidic residues are colored red and basic residues colored blue.

Interestingly, an inspection of the amino acid changes in StER3 (Table 3.3) showed that the polarity of the S1 subsite had been switched from acidic to basic. The electrostatic potential surface of the StER3 variant, modeled on the crystal structure of OmpT (14), is indicative of the reversal of polarity (change from red to blue color) at the bottom of the S1 subsite (Figure 3.5).

A



B

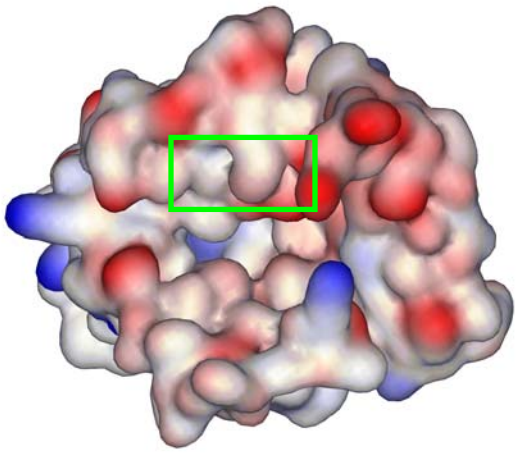


Figure 3.5: The electrostatic potential surfaces of (A) wild type OmpT and (B) StER3 variant rendered using WebLab ViewerLite (Accelrys, San Diego, CA). The green box indicates the putative S1 subsite. Red indicates regions of negative charge, blue indicates regions of positive charge and white patches represent hydrophobic segments.

In order to test if the altered specificity of the StER3 variant was indeed due to the remodeling of the S1 subsite, the three point mutants were constructed using the QuikChange protocol. Flow-cyometric evaluation of the Glu↓Arg activities of each of the point mutants demonstrated that none of them had any appreciable level of Glu↓Arg activity (Figure 3.6), indicating the synergistic nature of the three mutations.

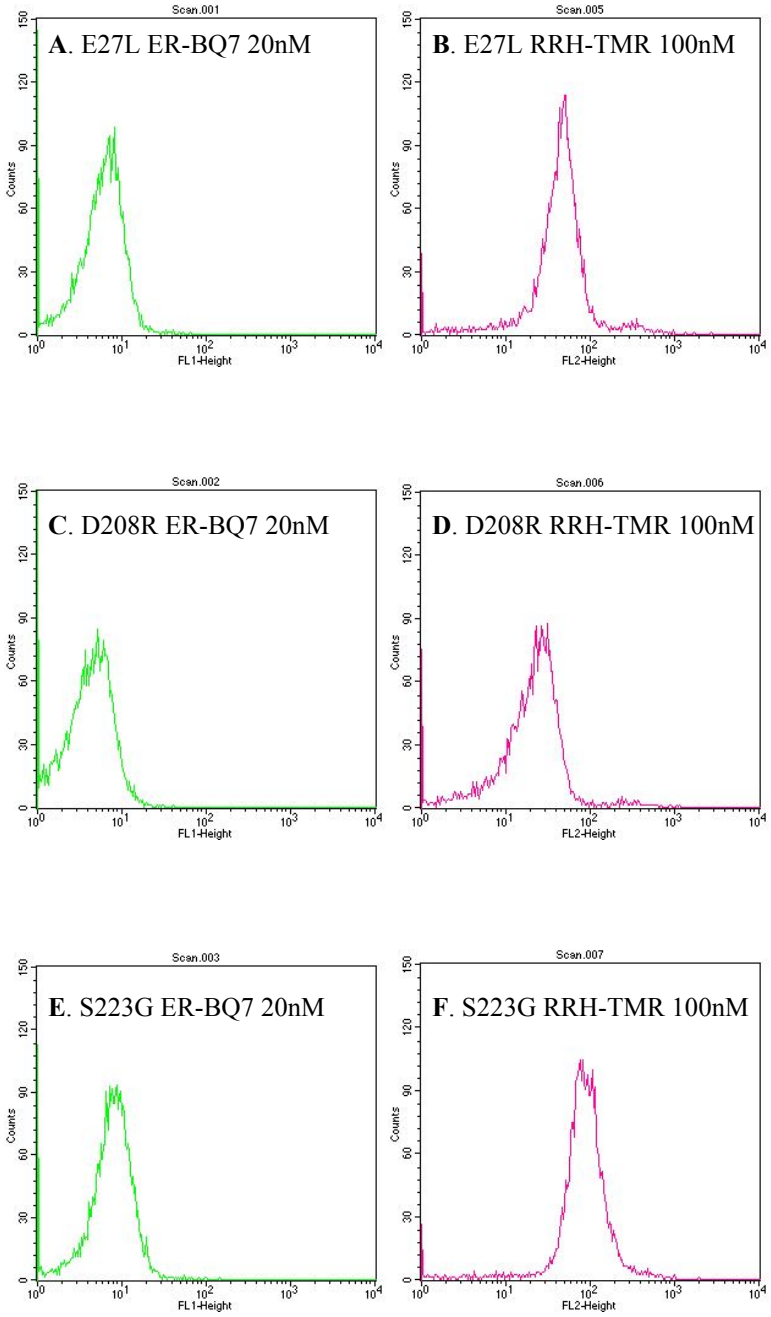


Figure 3.6: Fluorescence histograms of the three point mutants, (**A&B**) Glu27Val, (**C&D**) Asp208Arg and (**E&F**) Ser223Gly.

The protein was extracted from the membranes using *n*-octyl glucoside and purified (>90% as estimated by SDS-PAGE) (15) . In order to estimate the kinetic parameters, k_{cat} and K_M, 10-30 μM of the HPLC substrate, WCERVGKGRGR (Figure 3.1B), was incubated with 10 nM purified enzyme at 25^0C for 10 minutes and product formation was analyzed by monitoring tryptophan absorbance at 280 nm. The initial rates determined at a series of substrate concentrations (10-30 μM) were then fitted to a Michelis-Menten equation using non-linear regression algorithms. Kinetic analysis of StER3 confirmed that the selection yielded the expected outcome *i.e.* the engineering of a highly specific enzyme (Table 3.3). Additionally, the k_{cat}/K_M, of the protein with its preferred substrate, Glu↓Arg , was on par with the activity of OmpT with its preferred substrate, Arg↓Arg.

Enzyme	-Glu↓Arg-[1]	-Arg↓Arg-[2]	Specificity
	k_{cat}/K_M ($M^{-1}S^{-1}$)	k_{cat}/K_M($M^{-1}S^{-1}$)	-GluArg-/-Arg-Arg-
OmpT	----*	$2\pm1 \times 10^4$	------
StER3	$3\pm2 \times 10^5$	20 ± 10	1.5×10^4

Table 3.4: Kinetic parameters for the hydrolysis of unlabelled selection and counter-selection substrates by OmpT and StER3. Reactions were carried out at room temperature (25^oC). * No cleavage observed after a 24 h incubation of 30 μM substrate with 100 nM enzyme

[1] Substrate sequence WCERVGKGRGR

[2] Substrate sequence WEEGGRRVGKGR

Engineering of Tyr↓Arg activity: Having been successful in the isolation of the Glu↓Arg variant, StER3, the next target for the engineering of OmpT specificity, Tyr↓Arg, was chosen. The conversion of the P1 specificity of OmpT from Arg to Tyr is analogous to the conversion of the P1 specificity of trypsin to chymotrypsin (22). Screening of the Sat3 library, saturation at residues Glu27, Asp208 and Ser223, with Tyr↓Arg as the selection substrate and Arg↓Arg as the counter-selection substrate yielded no clones displaying an altered fluorescence profile. In order to design a more comprehensive library, all of the active-site residues potentially important for peptide binding/catalysis were first identified using a model of the OmpT crystal structure (14). Since Asp83, Asp85, Asp210 and His212 are believed to be critical for catalysis (14), these were excluded. The remaining 21 residues that constitute potential peptide binding/recognition sites were targeted for randomization. A library, 90NNS, was constructed by mixing randomized NNS primers with the primers for the wild-type codon, across these 21 residues (the actual residues are listed in the Materials and Methods section) in the assembly PCR reaction (Table 3.2) . The PCR fragment was re-amplified to enrich full length genes, gel-puirifed, digested and ligated into pDMLE19. The ligation was desalted and transformed into electrocompetent *E. coli* MC1061 and plated on selective media to yield $3x10^8$ transformants. DNA sequencing of 10 random clones from this library indicated that an average of 10 out of the 21 residues targeted were mutated in any given gene. The plasmid library isolated from these cells was retransformed into electrocompetent *E. coli* BL21(DE3); the cells grown up to an OD_{600}

of 2 units in selective media, washed with 1% sucrose and labeled with 20 nM Tyr↓Arg and 100 nM Arg↓Arg substrates. Flow-cytometric sorting was accomplished by setting gates based on FSC/SSC and FL1/FL2 profiles to enrich for cells expressing a Tyr↓Arg specific OmpT variant. After six rounds of sorting/resorting, the mean of the population remained constant. Five colonies from the final round of sorting were picked at random, grown up individually, labeled using the Tyr↓Arg and Arg↓Arg substrates, and their fluorescence profiles assayed by flow-cytometry. All five clones exhibited similar profiles in line with the screening criteria (Figure 3.7) and DNA sequencing revealed that all five clones encoded the same OmpT variant, 90YR3 (Table 3.4).

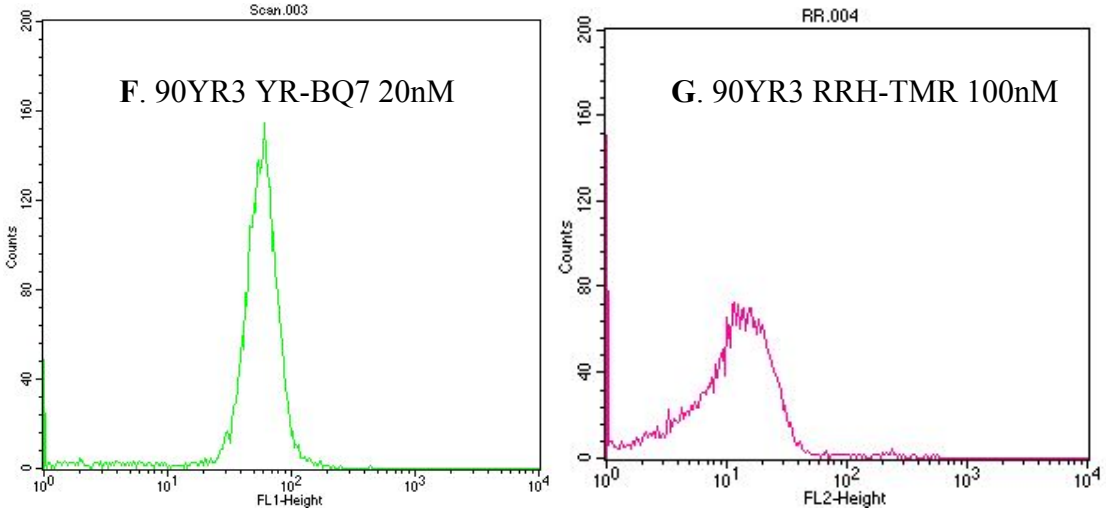


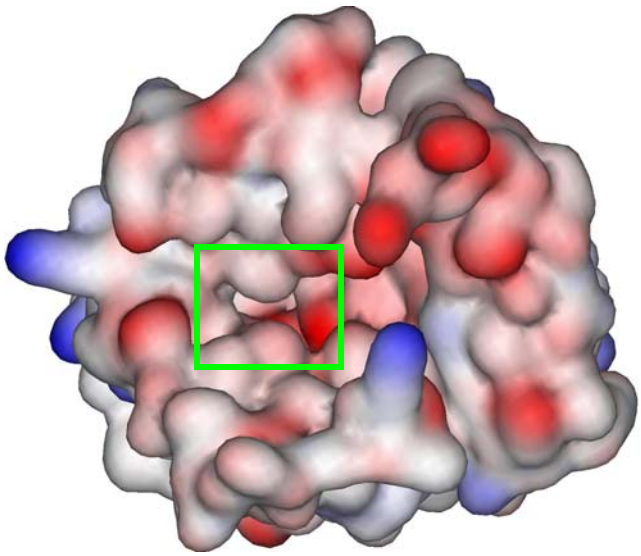
Figure 3.7: Fluorescence histograms of the 90YR3 variant labeled with the YR-BQ7 and RRH-TMR substrates.

Enzyme	Residue number						
	27	29	170	172	208	221	265
OmpT	Glu	Val	Ile	Tyr	Asp	Tyr	Leu
90YR3	Trp	Pro	Val	Val	His	Ala	Val

Table 3.5: List of amino acid changes in 90YR3 relative to OmpT.

Inspection of the amino acid changes in 90YR3 relative to OmpT indicated that aromatic residues had replaced the acidic residues at the bottom of the negatively charged S1 subsite (Glu27Trp & Asp208His). Since the Sat3 presumably library included these two residues, it is clear that although these residues are important for the altered specificity of 90YR3, they are not sufficient. An examination of the potential surface of 90YR3 in comparison to WT OmpT indicated an enlarged S1 subsite, which would be expected to better accommodate the bulkier tyrosine side chain (Figure 3.8)

A



B

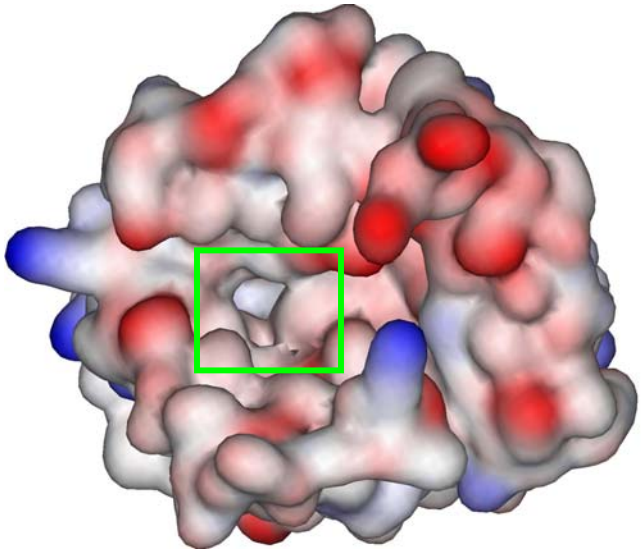


Figure 3.8: The potential surface of (A) OmpT & (B) 90YR3 generated using WebLab ViewerLite (Accelrys, San Diego, CA). The boxed region indicates the S1 subsite.

To estimate the kinetic parameters, k_{cat} and K_M, the enzyme was purified (>90% as estimated by SDS-PAGE) by extraction with *n*-octyl glucoside. Initial rates were determined by incubating 5 nMpurified enzyme with 10-30 μM HPLC substrate, WCYRVGKGRGR, in a 100 μl reaction at 25^{o}C for 15 minutes, and the initial rates were fitted to a Michaelis-Menten equation. The catalytic efficiency of 90YR3, k_{cat}/K_M with the Tyr↓Arg HPLC substrate was estimated as $8\pm3x10^4$, similar to the efficiency of OmpT with its preferred dibasic substrate. Additionally, the 90YR3 variant showed no cross-reactivity with the WT Arg-Arg substrate.

Engineering of Thr↓Arg activity: Initial attempts to isolate a highly active Thr↓Arg variant, based on screening the three residue saturation mutagenesis library, Sat3; the OmpT error-prone library, 1ePCR; and the 21-residue randomization library, 90NNS were unsuccessful. A focused library, which would allow a more complete radomization at fewer residues was sought. Accordingly, the seven residues that make up the S1 subsite, were identified based on the crystal structure of OmpT. A new library, designated Sat7 (Table 3.2), was constructed by using NNS primers encoding for these amino acids; 27,29,39,40,42, 208 and 223. Purification, digestion, ligation and subsequent transformation into *E. coli* MC1061 yielded $6x10^8$ independent transformants. Following plasmid isolation, re-transformation, growth and subsequent labeling, as above, a total of $2x10^7$ cells were analyzed on the MoFlo flow-cytometer with sorting gates set to collect 1% of cells displaying high FL-1 fluorescence (high activity with TR-BQ7) and low FL-2

fluorescence (low activity with RRH-TMR). Four rounds of sorting led to the isolation of the StTR2 variant. Fluorescence profiles of the StTR2 variant indicated that the enzyme was selective, but the overall fluorescence with the FACS selection substrate, Thr↓Arg indicated that the enzyme was unlikely to posses a high level of activity (Figure 3.9). Consistent with the flow-cytometric data, HPLC analysis of the cleavage of the Thr↓Arg substrate using purified enzyme (30-50 nM) showed very little cleavage (<10%) even after prolonged incubation (3 h).

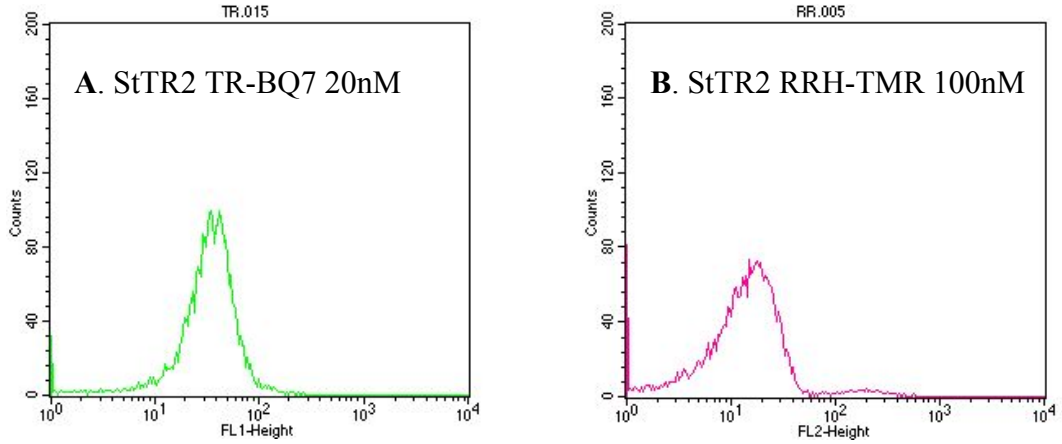


Figure 3.9: Fluorescence histograms of cells expressing the StTR2 variant labeled with the Thr↓Arg and Arg↓Arg FACS substrates.

In order to engineer an enzyme exhibiting more efficient Thr↓Arg hydrolysis, which in turn would result in higher fluorescence with the selection substrate, TR-BQ7 , the gene

encoding for StTR2 was subjected to random mutagenesis using error-prone PCR (Table 3.2). The resulting library was digested, ligated, and transformed into electrocompetent *E. coli* MC1061 cells to yield 2 x 10^8 independent transformants. DNA sequencing of 10 random clones from the library indicated an average of 8 substitutions per gene (0.8%). Library plasmid was pooled from transformed cells scraped from plates and grown for 5 generations. The pooled plasmids were re-transformed into *E. coli* BL21(DE3) ; the cells grown to an OD_{600} of 2 units, labeled using 20 nM Thr↓Arg and 100 nM Arg↓Arg and sorted. After three rounds of sorting, cells were sorted onto selective agar plates and grown at 37^{o}C overnight. Six clones were picked at random and their fluorescence profiles analyzed by flow-cytometry. All six clones showed high fluorescence when labeled with the TR-BQ7 substrate and low fluorescence when labeled with the RRH-TMR substrate (Figure 3.10). Sequencing revealed that all six clones encoded the same OmpT variant, eTR2 (Table 3.6).

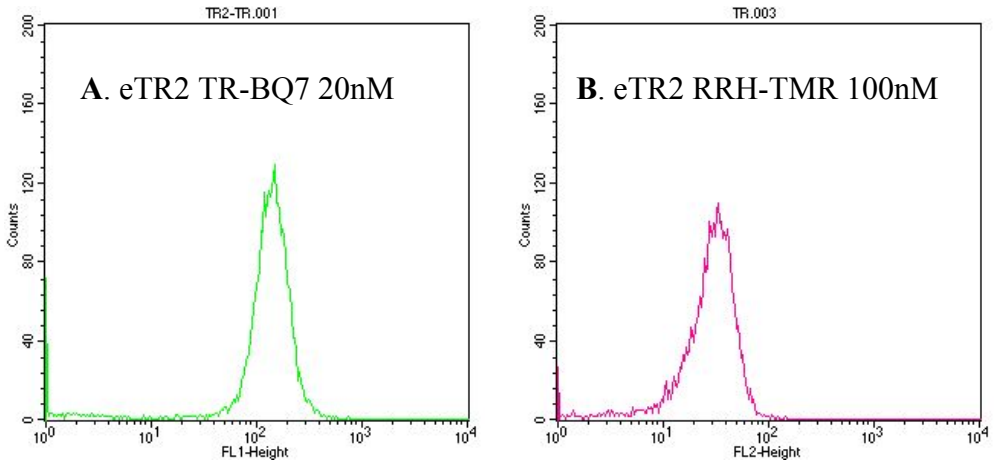


Figure 3.10: Fluorescence histograms of the eTR2 variant labeled with the TR-BQ7 (selection) and RRH-TMR (counter-selection) substrates.

A comparison of the amino acid sequences of both the StTR2 and eTR2 variants showed that they had four and nine changes, respectively, relative to WT OmpT (Table 3.6). Although seven amino acids (27,29,29,40,42,208,223) were targeted for randomization in the parent Sat7 library, the StTR2 variant had only four mutations; Glu27His, Val29Ser, Asp208Leu and Ser223Asp. The Val29Ser change at the entrance of the S1 subsite, can potentially hydrogen bond to the threonine in the substrate facilitating recognition. It is not obvious how the five additional changes in eTR2 increase Thr↓Arg activity.

Enzyme	Residue number								
	27	29	208	214	223	243	253	270	276
OmpT	Glu	Val	Asp	Asp	Ser	Pro	Trp	Asn	Ser
StTR2	*His*	*Ser*	*Leu*	Asp	*Asp*	Pro	Trp	Asn	Ser
eTR2	*His*	*Ser*	*Leu*	Asn	*Asp*	Ser	Gly	Tyr	Gly

Table 3.6: List of mutations in StTR2 and eTR2, relative to OmpT. The mutations in eTR2 derived from StTR2 (parent) are italicized.

The kinetic parameters for the hydrolysis of Thr↓Arg HPLC substrate were determined by incubating 15 nM of the purified enzyme (>80% pure SDS-PAGE) with 10-30 μM HPLC peptide, WCTRVGKGRGR, at 25^{o}C for 30 minutes. It is noteworthy that the catalytic efficiency, k_{cat}/K_M, of the eTR2 substrate (2±1 *10^4 $M^{-1}S^{-1}$) with its preferred substrate, Thr↓Arg was almost the same as that of OmpT with its preferred substrate containing an Arg↓Arg cleavage site. As always, mass spectrometry (LC-MS ESI analysis) was used to assign the cleavage site.

Engineering of Pro↓Arg activity: Despite screening four different libraries; the three residue saturation mutagenesis library, Sat3; the OmpT error-prone library, 1ePCR; the seven residue NNS library, Sat7, and the 21-residue randomization library, 90NNS,

using dual-color selection/counter-selection assays, OmpT variants that would selectively hydrolyze a Pro↓Arg substrate could not be isolated. The best enzyme, 90PR3, isolated by screening a 90NNS library, showed a moderate increase in fluorescence (Figure 3.11) compared to background (8-10 fold increase in fluorescence). Because the fluorescence of this variant labeled with the PR-BQ7 substrate was not indicative of a high activity clone, further characterization was not performed. In an attempt to isolate a variant capable of efficiently hydrolyzing Pro↓Arg sequences, an error-prone library was constructed based on the 90PR3 variant as the template. Gel-purification, restriction digestion, ligation into pDMLE19 and subsequent transformation into electrocompetent *E. coli* MC1061 yielded $3x10^8$ transformants. Sequencing revealed an average of 8 nucleotide changes per gene (0.8%). The plasmid library isolated from the pooled, scraped cells was retransformed into *E. coli* BL21(DE3) and the cells grown to an OD_{600} of 2 units. The cells were washed with 1% sucrose and an aliquot of the cells labeled using 20 nM Pro↓Arg and 100 nM Arg↓Arg. After the 10 minute incubation, $2.6x10^7$ cells were analyzed on the flow-cytometer and 1.2% of the total population was collected based on high FL-1 fluorescence and low FL-2 fluorescence. After four rounds of sorting, no enrichment was observed and hence no further sorting was done.

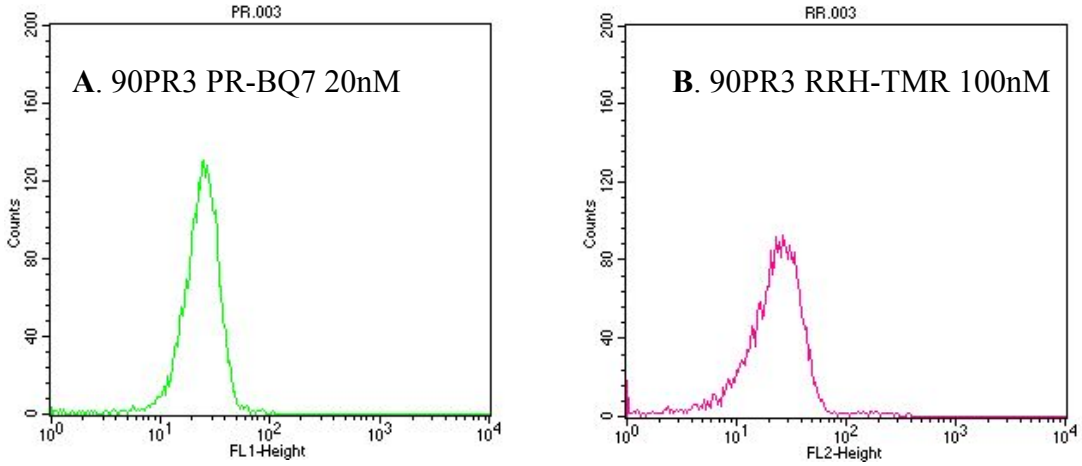


Figure 3.11: Fluorescence histograms of 90PR3 labeled with PR-BQ7 (selection substrate) and RRH-TMR (counter-selection substrate).

Sequencing of the genes encode by 90PR3 revealed that it had nine amino acid changes relative to OmpT. However, (Table 3.7) upon further analysis the enzyme was found not to be very specific and therefore was not further characterized.

Enzyme	Residue number								
	81	97	150	163	170	208	221	265	282
OmpT	Met	Asp	Tyr	Phe	Ile	Asp	Tyr	Leu	Ile
90PR3	Glu	Ser	His	Phe	Asp	Leu	Leu	Thr	Met

Table 3.7: List of amino acid changes in 90PR3 relative to OmpT.

Substrate selectivity of the engineered OmpT variants: As discussed above, enzyme kinetics analysis with peptide containing either the selection or the counterselection peptide revealed that the engineered OmpT variants were highly selective. We then examined whether the isolated enzyme variants could cleave sequences other than the selection peptide. Initial conditions for cleavage were determined using 1-20 nM purified enzymes with 30μM selection substrate such that the enzyme cleaved 40-50% of the selection substrate in 45 minutes. After having determined the optimal initial conditions, the enzyme concentration was set and the percentage cleavage determined for a panel of 6 substrates that have the P1 amino acid systematically changed to include one amino acid from the six classes of amino acids (Table 3.12) (18). WT OmpT prefers the dibasic peptide and shows a weak activity for cleavage at Ala↓Arg (80-fold lower k_{cat}/K_M, see chapter 2), but failed to cleave any of the other peptides at Xaa↓Arg (There were secondary cleavage sites on the peptides, please see table 3.12). The OmpTS223Arg mutant and StER3 show no cross-reactivity with any of the other substrates, indicating that they were remarkably specific enzymes. 90YR3, the enzyme selected for high activity towards Tyr↓Arg, shows a weak cross-reactivity for cleavage at Ala↓Arg sequences but is otherwise specific. The eTR2 variant shows cross-reactivity for cleavage at Glu↓Arg sequences and weak activity for cleavage at Ala↓Arg but it preferentially cleaves Thr↓Arg sequences, consistent with its selection criterion. The systematic analysis of the P1 specificity of these variants demonstrates that these genuinely enzymes are selective even when assayed with a diverse set of substrates.

Enzyme	Number of mutations	Relative substrate preferences						k_{cat}/K_M ($M^{-1}s^{-1}$)
		R↓R[1]	A↓R[2]	E↓R[2]	Y↓R[2]	T↓R[2]	P↓R[2]	
WTOmpT	-	1	0.04	-	-[#]	-*	-*	$2\pm1 \times 10^{4}$ [&]
S223R	1	-	1	-	-	-	-	$2.6\pm0.8 \times 10^{5}$
StER3	3	-	-	1	-	-	-	$3\pm2 \times 10^{5}$
90YR3	7	-	0.2	-	1	-	-	$8\pm3 \times 10^{4}$
eTR2	9	-	0.05	0.3	-	1	-	$2\pm1 \times 10^{4}$

Table 3.8: Kinetic characterization and the relative substrate preferences of wild type OmpT and the OmpT variants measured at room temperature (25°C). A dash (-) indicates that no substrate cleavage was observed under the conditions.

[1] Substrate sequence: WEEGGRRVGKGR

[2] Substrate sequence: WCXRVGKGRGR where X = A/E/Y/T/P

[#] OmpT cleaves the peptide between R↓V

* OmpT cleaves the peptide between K↓G

[&] OmpT cleaves the peptide sequence WEEGGR↓RIGRGGK with k_{cat}/K_M of $1.7\pm0.4*10^{5}$ (Chapter 2).

3.3.2 Engineering the P1' specificity of OmpT

Having successfully remodeled the S1 substrate binding pocket to recognize a diverse set of amino acids, the next step was to explore changing the selectivity for the S1' subsite.

Engineering of Arg↓Val activity: Olsen et al. had previously reported the engineering of an OmpT variant C5 that showed a 60- fold improvement towards the hydrolysis of Arg↓Val substrates (23) compared to WT OmpT. That enzyme, however, turned out to be non-specific and cleaved the peptide sequence, WGGPGRVVGGTI, at multiple sites. OmpT showed weak activity for cleavage at Arg↓Val with the Tyr-Arg-Val substrate (Table 3.12). The Tyr-Arg-Val substrate was used as the selection substrate in an attempt to engineer a selective Arg↓Val variant. Also, the counter-selection substrate has the Arg-Arg-Val sequence (Figure 3.2) and counter-selection using this substrate while selecting with the Tyr-Arg-Val substrate would facilitate isolation of variants that cleave Tyr↓Arg-Val sequences and not Arg↓Val sequences. Hence, library screening was performed using selection-only assays.

The plasmid library isolated from the three residue saturation library, Sat3, was freshly transformed into electrocompetent *E. coli* BL21(DE3) and grown up to an OD_{600} of 2 units. The cells were washed and resuspended in 1% sucrose and labeled with 20 nM Tyr-Arg↓Val substrate. After a ten-minute incubation, the library was sorted on the MoFlo and gates were set to collect 1% of the most fluorescent cells. Following four

rounds of sorting, the cells were sorted onto LB/ampicillin (200μg/ml) plates. Five randomly selected clones when assayed on the flow-cytometer showed identical fluorescence profiles (Figure 3.12). Surprisingly, although three residues were targeted for randomization, the best variant, encoded by all five clones, had only one amino acid change, OmpTSer223Asp. Not surprisingly, the OmpTSer223Asp mutation increases the overall negative charge in concert with the other two WT residues, Glu27 and Asp208 and may help better recognize the P1 arginine on the substrate. Initial kinetic characterization using 50 nM purified enzyme and 30 μM substrate indicated that this was not a high activity enzyme and hence further characterization was not performed.

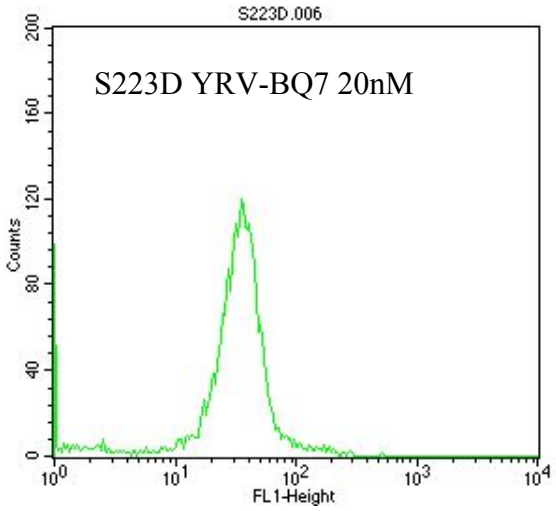


Figure 3.12: Fluorescence histogram of the OmpTSer223Asp variant labeled with the Tyr-Arg↓Val substrate.

To engineer a variant that showed high activity towards the hydrolysis of Arg↓Val sequences, an error-prone library was constructed using *ompTSer223Asp* gene as the template in mutagenic PCR. The resulting library, S223DePCR, was gel-purified, restriction enzyme digested, ligated into pDUCE19 and transformed into electrocompetent *E. coli* MC1061 to yield $2x10^8$ transformants. The plasmid library was isolated from pooled, scraped cells; retransformed into *E. coli* BL21(DE3) and grown up in 1 l of 2xYT containing 200 μg/ml to an OD_{600} of 2 units. The cells were washed with 1% sucrose and labeled with 20 nM of Tyr-Arg↓Val substrate and analyzed on the flow-cytometer. Not surprisingly, the majority of the clones displayed a fluorescence profile similar to OmpTSer223Asp. The top 1% fluorescent cells were sorted into 2xYT, grown up, labeled, and resorted. After three rounds of sorting, the cells were sorted onto LB/ampicllin plates. Cells expressing the consensus variant, eRV6, showed bright green fluorescence, indicating the presence of a highly active enzyme (Figure 3.13).

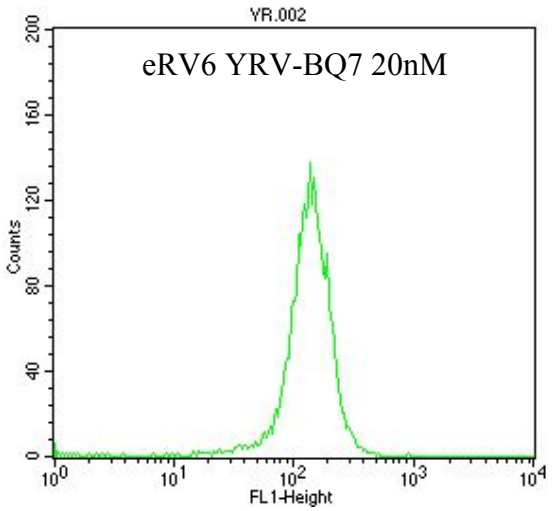


Figure 3.13: Fluorescence histogram of the eRV6 variant labeled with the Tyr-Arg↓Val substrate.

Sequencing of the gene indicated that it contained 2 nucleotide changes resulting in two amino acid substitutions; Gln63Arg and Asp97His. In addition, it contained the Ser223Asp substitution found in the parent clone used as the template for error-prone PCR (Table 3.9). The kinetic parameters for the hydrolysis of Arg↓Val HPLC substrate were determined by incubating 1 nM of the purified enzyme (>90% pure SDS-PAGE) with 10-30 μM HPLC peptide, WCYRVGKGRGR, at 25°C for 10 minutes. Kinetic analysis of the eRV6 variant confirmed that the enzyme was highly active with a k_{cat}/K_M of 5±2 x10^5 $M^{-1}s^{-1}$ and mass spectrometry was used to assign the cleavage between Arg↓Val (Figure 3.13)

Enzyme	Residue number		
	63	97	223
OmpT	Gln	Asp	Ser
eRV6	Arg	His	Asp

Table 3.9: List of amino acid changes in eRV6 relative to OmpT.

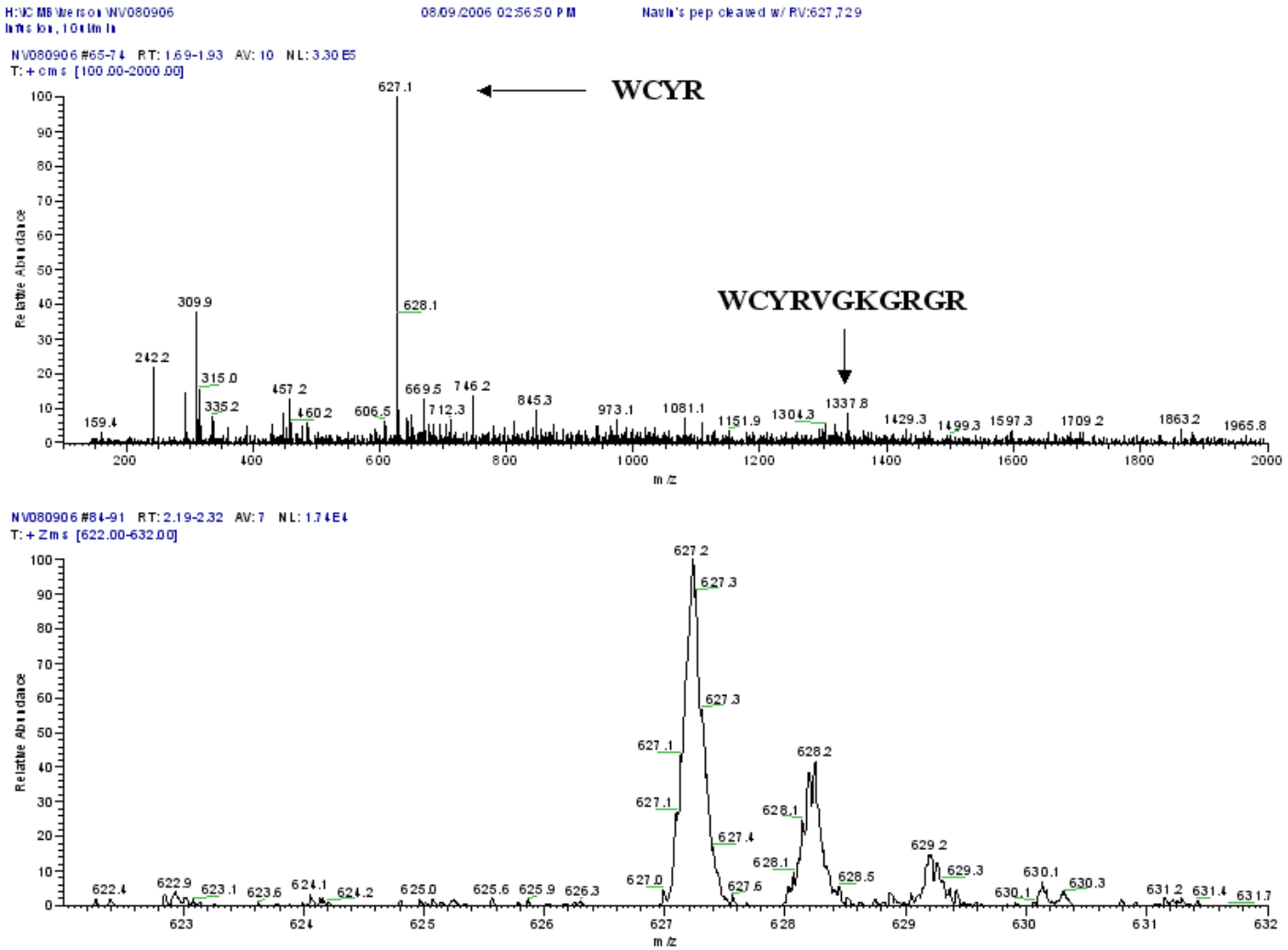
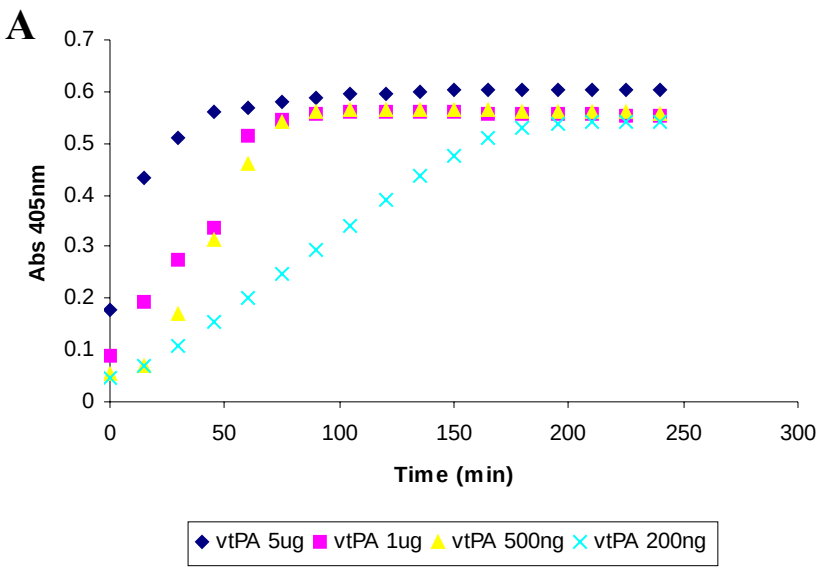


Figure 3.14: LC-MS ESI of the cleavage product of eRV6 with the Ty-Arg↓Val substrate.

Plasminogen activation assay: The site-specific activation of plasminogen to plasmin occurs by cleavage between Arg561↓Val562. In order to investigate the ability of the engineered OmpT Arg↓Val mutant, eRV6, to activate plasminogen, the enzyme was purified and incubated with human-Glu-type plasminogen. Plasmin formation analyzed by virtue of its ability to hydrolyze a chromogenic substrate in a coupled assay (24). The direct hydrolysis of the chromogenic substrate by eRV6 in the absence of plasminogen was employed as a negative control, to rule out any unwanted activity towards the chromogenic substrate. The OmpT mutant, eRV6, showed a concentration dependant activation of plasminogen to plasmin (Figure 3.14) and showed no increase in absorbance in the absence of plasminogen (not shown). OmpT, consistent with earlier reports (25), showed only weak plasminogen activation activity.



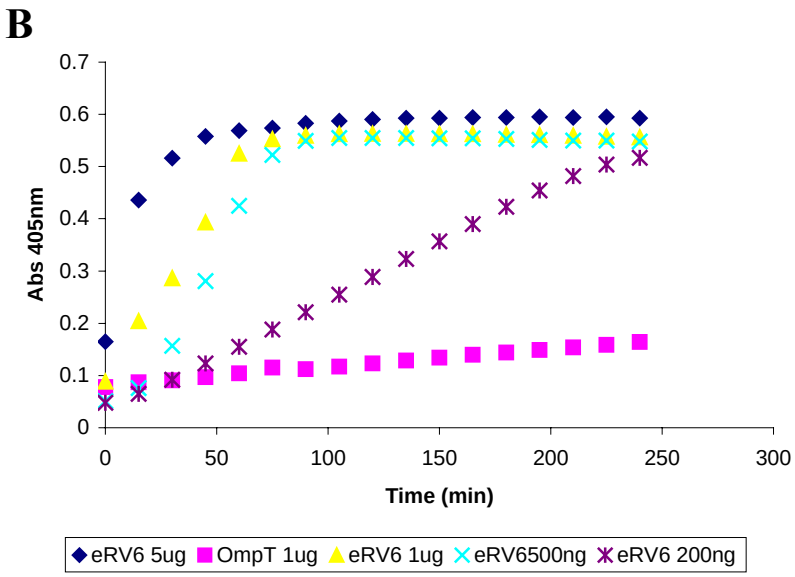


Figure 3.15: Concentration dependant activation of plasminogen by (A) purified tPA and (B) eRV6 & OmpT

Engineering of Glu↓Ala activity: As mentioned before, WT OmpT preferentially cleaves at dibasic residues (Arg preferred in both P1 and P1'). Having successfully demonstrated the engineering of OmpT to accept diverse substrates in S1, the ability to isolate OmpT variants that selectively hydrolyzed unrelated, non-preferred sequences like Glu-Ala was explored. The DNA encoding the Glu↓Arg mutant, StER3, was used as a template to construct two different libraries; the first library involved saturation mutagenesis of residues 87, 97 and 150; and the second library was constructed by error-prone PCR (Table 3.2). Again, following cloning and transformation into electrocompetent *E. coli* MC1061, the error-prone library, ER3ePCR yielded $1x10^8$ transformants. Sequencing ten random clones revealed that the average mutation rate was 8 nucleotides per gene (0.8%). Since a much smaller library was sufficient to cover the diversity of the three-residue saturation library, the ligation was directly transformed into electrocompetent *E. coli* BL21(DE3). The two libraries, ER3PCR and ER3Sat3, were grown up to an OD_{600} of 2 units in 1 l of 2xYT containing 200 μg/ml ampicillin at 37^oC. The cells were washed with 1% sucrose and labeled with 20 nM Glu↓Arg for ten minutes. An aliquot of the labeling reactions were analyzed on the flow-cytometer and gates set based on FSC/SSC and FL-1 to sort 0.9% - 1.1% of the most fluorescent cells. After four rounds of sorting, no enrichment was observed with the ER3Sat3 library. On the other hand, the ER3ePCR library after five rounds of sorting showed promising enrichment and cells were sorted onto selective plates. Five clones picked at random showed similar fluorescence profiles when assayed by flow-cytometry (Figure 3.15).

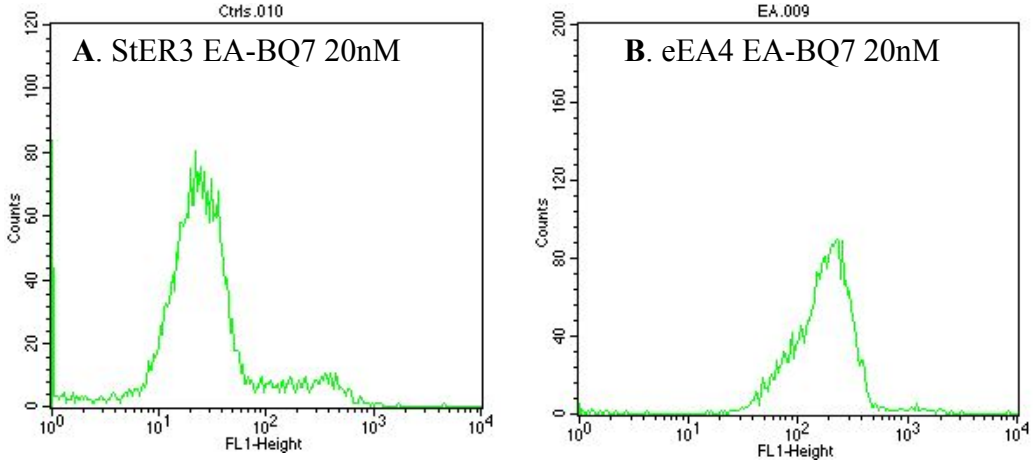


Figure 3.16: Fluorescence histograms of (A) StER3 & (B) eEA4 labeled with the Glu↓Ala substrate.

Sequencing of the gene encoding the five different clones led to the identification of the consensus eEA4 variant that had seven additional changes compared to StER3, the starting parent (Table 3.10)

Enzyme	Residue number									
	27	48	183	208	214	216	223	244	274	280
OmpT	Glu	Asn	Leu	Asp	Asp	Gly	Ser	Asn	Asp	Ala
StER3	Leu	Asn	Leu	Arg	Asp	Gly	Gly	Asn	Asp	Ala
eEA4	*Leu*	Asp	Phe	*Arg*	Asn	Glu	*Gly*	Ile	Gly	Glu

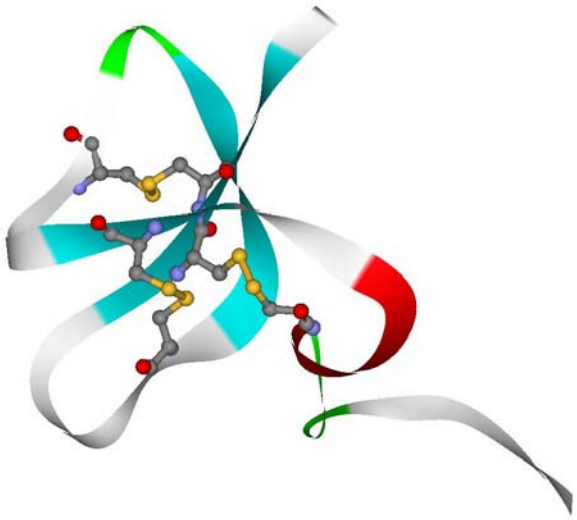
Table 3.10: Amino acid changes in eEA4 relative to OmpT & StER3. The changes common to StER3 and eEA4 are italicized.

In order to determine the kinetic parameters of the eEA4 variant towards hydrolysis of the Glu↓Ala substrate, 6-20 μM of the HPLC substrate, WCEAVGKGRGR, was incubated with 1 nM purified enzyme (>90% as estimated by SDS-PAGE) at 25^{o}C for 10min. The measured catalytic efficiency, k_{cat}/K_M, of the eEA4 variant was $1\pm0.3\text{x}10^6$, better than the efficiency of OmpT with its preferred Arg-Arg substrate, $2\pm1\text{x}10^4$. Cleavage at Glu↓Ala by the eEA4 variant was confirmed using mass-spectrometry.

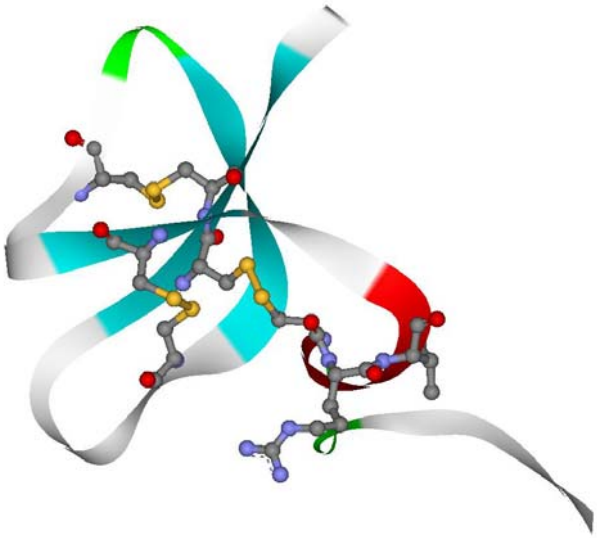
3.3.3 Human-β-defensin3 assays: The systematic engineering of substrate specificity demonstrated that OmpT is a surprisingly plastic enzyme, capable of adapting to a wide variety of new substrates. The ability of OmpT to cleave highly cationic antimicrobial peptides has been reported before (26). The co-evolution of cationic antimicrobial defense peptides and the counter resistance mechanisms in bacteria represents a fine

example of symbiotic adaptive evolution (27). β-Defensins are cationic antimicrobials that consist of compactly folded β-sheets held together by multiple disulfide bonds that renders them resistant to bacterial proteolysis. In order to test the hypothesis that the plasticity of OmpT might contribute towards facile adaptation to cleave a changing array of antimicrobials, the potential cleavage sites for the engineered OmpT variants in human-β-defensin3 (HBD-3) were mapped out (Figure 3.16). Two potential sites within HBD-3 could be cleaved by mutants generated as part of this study;, Arg12↓Val13 is a putative substarte foreRV6 and Thr35r↓Arg36 for eTR2. The bacteriocidal effect of HBD-3 towards *E. coli* BL21(DE3) [*ompT, ompP*] cells expressing no enzyme; WT OmpT, eRV6, or eTR2 was determined using the liquid microdilution assay (17). Briefly, ~10^7 cells were incubated with 40 μg/ml HBD-3/water (control) in salt-free buffer at 37C for 3h and then serial dilutions were plated to estimate viability (Table 3.11). Cells expressing OmpT showed a 33-fold higher viability compared to cells with no enzyme. Importantly the cells expressing either eRV6 or eTR2 showed >2000 fold improvement in viability.

A



B



C

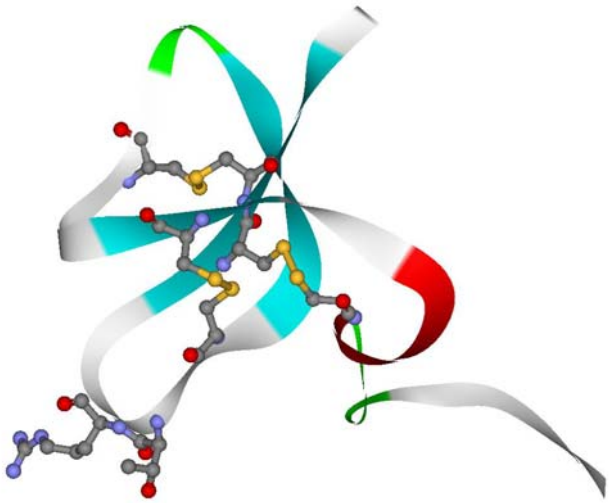


Figure 3.17: Solution structures of (A) human- β-defensin3 with the three disulfide bridges (HBD-3, PDB ID: 1KJ6) (B) HBD-3 with Arg12 & Val13 residues highlighted & (C) HBD-3 with the Thr35 & Arg36 residues highlighted.

Cells	HBD-3concentration (ug/ml)	Cell count [#]	% Viability
BL21(DE3)	- 40	2×10^7 $<10^3$	< 0.006
BL21(DE3)-OmpT	- 40	1.2×10^7 2×10^4	0.2
BL21(DE3)-eRV6	- 40	6×10^6 1×10^6	17
BL21(DE3)-eTR2	- 40	6×10^6 8×10^5	13

Table 3.11: Cell viability assays measured in presence and absence of human-β -defensin3. Briefly, ~1* 10^7 cells were incubated with HBD-3 at a concentration of 40 µg/ml in 10 mM Na_2HPO_4, pH=7.4 at 37°C for 3 h and the viability as a percentage of the same cells incubated with water as a control is reported. [#] The standard error in these assays is ±25%.

Cleavage of T7 RNA polymerase and auto-proteolysis (preliminary data): The ability of the engineered OmpT variants to cleave folded proteins was tested using T7 RNA polymerase. The ability of OmpT to cleave T7 RNA polymerase (Molecular weight ~100 kDa) between Lys172↓Arg173 has been reported before(28). 300U of T7 RNA polymerase was incubated with 50ng of OmpT at 37^{o}C for 2h. The reaction was then run on a 8-16 % gradient Tris-glycine gel and stained using GelCodeBlue ®. Four different bands were observed; 80 kDa, 55k Da, 25 kDa and 20 kDa. N-terminal sequencing was used to assign the cleavage sites for the two larger bands; 80 kDa Lys172↓Arg173, and 55 kDa Arg391↓Lys392. When 300 units of T7 RNA polymerase was similarly incubated with 50-200 ng of the OmpT variants (OmpTS223R, StER3, eTR2, eRV6, eEA4) at 37^{o}C for 2 h, no cleavage was observed.

OmpT undergoes auto-proteolysis by self-cleavage between Lys217↓Arg218. In order to assess the ability of the engineered OmpT variants to self-cleave, reactions were setup with 5 µg of OmpT/OmpT variants in 10 mM MES, 10 mM EDTA, pH = 6.1 at 37^{o}C overnight. The reaction was run on a 4-20 % Tris-glycine gradient gel and stained using GelCodeBlue ®. WT OmpT showed two bands; 25 kDa (Ser1-Lys217) and 9 kDa (faint band, Arg218-Phe297). All of the engineered OmpT variants (OmpTS223R, StER3, eTR2, eRV6, eEA4) showed a single band (~34 kDa) corresponding to intact protein.

3.4 Discussion

We have engineered the substrate specificity of the *E. coli* outer membrane protease OmpT from large libraries screened quantitatively for activity and selectivity towards multiple new substrates using our novel dual-color selection and counter-selection flow-cytometric high-throughput assay. Because OmpT is not folded and active until it reaches the outer membrane (13), and since the catalytic core is directed towards the external milieu (14), the protease does not cleave essential physiological proteins or itself, during *in vivo* overexpression. Therefore, protease libraries with diverse specificities can be expressed without killing the host cell. Additionally, OmpT has many properties desirable for practical applications such as high thermostability, and activity in high concentrations of denaturants (29) or reducing agents such as DTT/TCEP (OmpT does not contain any disulfides that could be subject to reduction by such agents).

To enable the isolation of OmpT variants that recognize substrates with altered P1 specificities, a set of libraries were constructed using PCR based gene assembly and random mutagenesis. The first target for the engineering of P1 specificity of OmpT, Glu↓Arg, required the electrostatic reversal of ion-pair specificity (swapping positively and negatively charged residues between the substrate and the substrate binding-pocket in the enzyme). Protein engineering has previously been used to explore the reversal of charges between the substrate and the substrate binding pocket in many enzymes such as trypsin and aspartate aminotransferase (19, 20). However, these efforts led to non-specific enzyme variants that were not highly active. Because the microenvironment in

the substrate binding pocket in the WT enzyme had likely been optimized in the context of the native residue/substrate, introduction of the opposite charge would completely change the dielectric constant of the pocket (21). For this reason, electrostatic reversal of ion-pair specificity was deemed undoable (21). To overcome this problem, we sought to remodel all the residues that comprised the bottom of the S1 binding pocket in OmpT. Accordingly, a saturation library, Sat3, that resulted in the randomization of Glu27, Asp208 and Ser223 was constructed and the high-throughput dual-color selection assay was used to screen for OmpT variants that could selectively recognize Glu↓Arg substrate (Glu in P1). Multiple rounds of sorting led to the isolation of a highly active and specific variant, StER3. Flow cytometric analysis and *in vitro* kinetic characterization confirmed the alteration of specificity of StER3 towards Glu↓Arg substrates. Additionally, the purified StER3 variant showed extremely high activity with its preferred Glu↓Arg substrate, comparable to that of OmpT with its preferred dibasic substrate, Arg↓Arg. Construction and subsequent flow-cytometric analysis of the three individual point mutations that constitute StER3, suggested that these mutations are epistatic with respect to conferring Glu↓Arg activity, that is they work in a synergistic fashion that is not simply additive.

Engineering the conversion of OmpT to recognize tyrosine in P1 is directly analogous to the conversion of the P1 specificity of trypsin to chymotrypsin (22). Chymotrypsin and trypsin are serine proteases that have a high degree of both sequence and structure homology. Chymotrypsin cleaves after large hydrophobic residues while trypsin cleaves

after basic residues. In spite of this homology, conversion of trypsin into a chymotrypsin like protease (preference for large hydrophobic residues) by replacement of the S1 site residues with the corresponding ones in chymotrypsin was not successful (22). The conversion of trypsin to chymotrypsin was achieved by grafting entire surface loops from chymotrypsin (L1, L2), that were shown to be critical for amide hydrolysis rather than substrate recognition/binding, into trypsin.

The construction of a semi-random mutagenesis library that targeted all of the putative peptide contact-residues (excluding the four residues, Asp83, Asp85, Asp210 and His212 important for catalysis), 90NNS, and subsequent screening with the Tyr↓Arg and Arg↓Arg substrates, allowed the isolation of the OmpT variant 90YR3. Again, thorough characterization of 90YR3 demonstrated that altered specificity was not engineered at the cost of high overall activity. Nine amino acid changes in OmpT were sufficient to change it from a trypsin like protease to a chymotrypsin like protease (P1 specificity).

Because attempts to isolate a highly active, specific Thr↓Arg variant by screening multiple libraries was unsuccessful; a two-step approach was used. First, a medium activity but specific variant, StTR2 was isolated from a Sat7 library. Next, construction of an error-prone library based on StTR2 and screening using Thr↓Arg & Arg↓Arg substrates facilitated the engineering of a highly active variant, eTR2.

In order to truly understand the selectivity of these engineered protease variants, their cross-reactivity was determined across all classes of amino acids preferred/tolerated in P1. OmpT, consistent with earlier reports, preferred basic residues in P1 and showed no

cleavage after Glu/Tyr/Thr/Pro, although Ala was tolerated (13). Two enzymes, OmpTS223R and StER3, showed no cross-reactivity with substrates other than their respective selection substrates. The chymotrypsin like variant, 90YR3, showed no preference for substrates containing Glu/Thr/Pro/Arg in P1, although the Ala residue was tolerated with a lower level of activity compared to Tyr. The high activity Thr↓Arg variant, eTR2, showed no cleavage at Arg/Pro/Tyr, but did show appreciable cleavage after Glu and a weak ability to hydrolyze after Ala. However, the preferred residue was clearly Thr in P1, consistent with the way this enzyme was selected on the basis of high activity for cleavage of the Thr↓Arg substrate. It has been previously postulated that "moonlighting activities" (activities not under direct selection) pop up stochastically in selection only environments (leading to non-specific enzymes) and that the moonlighting activity may confer a phenotypic advantage (30, 31). Consistent with other directed evolution experiments (32), the presence of an explicit negative constraint seems to eliminate non-specific generalist enzymes and commits the enzyme pool towards variants that have the desired high levels of activity and specificity towards the selection substrate.

Since the isolation of OmpT variants with altered P1 specificities was successful, the engineering of the P1' specificity of OmpT was attempted. Again, a two-step approach was used to engineer an OmpT variant that showed high-levels of activity with Arg↓Val substrate. Flow-cytometry and *in vitro* kinetic characterization confirmed that the eRV6

variant was highly active for the hydrolysis at Arg↓Val sequences. The eRV6 variant, surprisingly, had only three mutations relative to OmpT.

As part of the blood-clotting cascade fibrin-bound tissue plasminogen activators like tPA cleave plasminogen to yield a serine protease plasmin which then dissolves the cross-linked fibrin clots into soluble fragments. The activity of plasmin is tightly regulated by plasmin inhibtor serpin α2-antiplasmin (α2AP). The Pla protein of *Yersinia pestis* is a 292 amino acid outer membrane protease that shows about a 50% homology at the amino acid level to OmpT. Pla, an essential protein for the invasive character of plague (33), can activate plasminogen and degrade the α2AP leading to unregulated proteolysis by plasmin (34). The site-specific activation of plasminogen to plasmin occurs by cleavage between Arg561↓Val562. To investigate the ability of our engineered OmpT Arg↓Val mutant eRV6 to activate plasminogen, the purified enzyme was incubated with human-Glu-type plasminogen and plasmin formation analyzed by virtue of its ability to hydrolyze a chromogenic substrate. The enzyme showed a concentration dependent activation of plasminogen, similar to vtPA, while OmpT, consistent with earlier reports (25), showed only weak plasminogen activation. Interestingly, a sequence alignment of Pla and the eRV6 amino acid sequences indicates that the Ser223Asp mutation of eRV6 is homologous to Glu217 of Pla (Figure 3.17). Importantly, the conversion OmpT into a plasminogen activator has previously been achieved through exchange of loops between OmpT and Pla (34). Through the application of a flow-cytometric high-throughput sassay

to screen large libraries, the same phenotype was isolated that required only three changes.

```
OmpT -STETLS-FTPDNINADISLGTLSGKTKERVYLAEEGGRKVSQLDWKFNNAAIIKGAINW 58
Pla  ASSQLIPNISPDSFTVAASTGMLSGKSHEMLYDAETG-RKISQLDWKIKNVAILKGDISW 59
eRV6 -STETLS-FTPDNINADISLGTLSGKTKERVYLAEEGGRKVSQLDWKFNNAAIIKGAINW 58
      *:: :. ::**.:..  * * ****::* :* ** * **:******::*.**:** *.*

OmpT DLMPQISIGAAGWTTLGSRGGNMVDQDWMDSSNPGTWTDESRHPDTQLNYANEFDLNIKG 118
Pla  DPYSFLTLNARGWTSLASGSGNMDDYDWMNE-NQSEWTDHSSHPATNVNHANEYDLNVKG 118
eRV6 DLMPQISIGAAGWTTLGSRGGNMVDQDWMDSSNPGTWTHESRHPDTQLNYANEFDLNIKG 118
     *  . :::.* ***:*.* .*** * ***:. * . **..* ** *::*:***:***:**

OmpT WLLNEPNYRLGLMAGYQESRYSFTARGGSYIYSSEEGFRDDIGSFPNGERAIGYKQRFKM 178
Pla  WLLQDENYKAGITAGYQETRFSWTATGGSYSYNN----GAYTGNFPKGVRVIGYNQRFSM 174
eRV6 WLLNEPNYRLGLMAGYQESRYSFTARGGSYIYSSEEGFRDDIGSFPNGERAIGYKQRFKM 178
     ***:: **: *: *****:*:*:** **** *..       *.**:* *.***:***.*

OmpT PYIGLTGSYRYEDFELGGTFKYSGWVESSDNDEHYDPGKRITYRSKVKDQNYYSVAVNAG 238
Pla  PYIGLAGQYRINDFELNALFKFSDWVRAHDNDEHY--MRDLTFREKTSGSRYYGTVINAG 232
eRV6 PYIGLTGSYRYEDFELGGTFKYSGWVESSDNDEHYDPGKRITYRDKVKDQNYYSVAVNAG 238
     *****:*.** :****.. **:*.**.: ******   : :*:*:*.....**...:***
```

```
OmpT YYVTPNAKVYVEGAWNRVTNKKGNTSLYDHNN-NTSDYSKNGAGIENYNFITTAGLKYTF 297
Pla  YYVTPNAKVFAEFTYSKYDEGKGGTQTIDKNSGDSVSIGGDAAGISNKNYTVTAGLQYRF 292
eRV6 YYVTPNAKVYVEGAWNRVTNKKGNTSLYDHNN-NTSDYSKNGAGIENYNFITTAGLKYTF 297
     *********:.* ::.:  : **.*.  *:*. :: . . :.***.* *: .****:* *
```

Figure 3.18: Sequence alignment of OmpT, Pla and eRV6 generated using ClustalW (http://www.ebi.ac.uk/clustalw/).

In order to completely rid the ariginine requirement of OmpT in both P1 and P1', the engineering of Glu↓Ala variant, starting with the Glu↓Arg enzyme, StER3, was attempted. The construction and screening of an error-prone library resulted in the isolation of a highly active Glu↓Ala variant, eEA4. The *in vitro* catalytic efficiency of the enzyme towards hydrolysis of the Glu↓Ala substrate was higher than the efficiency of OmpT with its preferred Arg↓Arg peptide. It is surprising that ten amino acid changes in the OmpT framework are enough to alter its specificity from Arg↓Arg to Glu↓Ala.

Since the cleavage of highly cationic antimicrobial peptides by OmpT has been reported before (26), the ability of the engineered OmpT variants reported here, to cleave more complex folded antimicrobials like defensins was explored. Not surprisingly, the two mutants that had recognition sites in human-β -defensin3, eTR2 and eRV6, when

overexpressed in cells showed a greater than 2000 fold increase in viability compared to cells expressing no enzyme. A particularly interesting aspect of this result is that β-defensins are believed to have evolved to resist proeteolysis (27) and yet overexpression of variants of the surface-displayed protease undermines its efficacy.

3.5 Conclusions

The ability to systematically modify the sequence specificity of OmpT to recognize unrelated, non-preferred sequences through the construction of large mutagenic libraries and high-throughput flow-cytometric screening has been demonstrated (Table 3.12). The activity of these engineered enzymes with their preferred substrates was as good (an in certain cases better) as OmpT with its preferred dibasic substrate. An examination of the overall specificity of these engineered OmpT variants demonstrated that high overall activity with the selection substrates was not achieved at the cost of overall specificity. This represents one of the first examples of engineered enzymes that posses both high overall activity and specificity (See Table 1.1 Chapter 1 and Table 3.12). The ability to engineer high activity with novel substrates, for which the starting parent has no activity, is particularly rare (35) (See also Table 1.1 Chapter 1 and Table 3.12).

The systematic protease engineering shown here opens up other questions that can be answered fairly easily. For example, is it possible to recombine variants that recognize altered P1 amino acids with variants that recognize altered P1' amino acids either through site-directed mutagenesis or using DNA shuffling and subsequent screening to create

chimeric enzymes that recognize a hybrid substrate sequence? Is it possible to counter-select against multiple sequences and still isolate highly active enzymes? These questions are currently being explored through the implementation of high-throughput flow-cytometric assays.

Parental Enzyme(s)	WT substrate/ New substrate Specific for new	WT enzyme k_{cat}/K_M ($M^{-1}s^{-1}$)	Best Enzyme variant k_{cat}/K_M	Comments. New activity present in parent enzy`me (Y/N)
OmpT	Arg↓Arg / Ala↓Arg (Y)	$1.7\pm0.8*10^5$	$2.6\pm0.8*10^5$	100-fold improvement in activity; 3 million-fold change in specificity (Y)
OmpT	Arg↓Arg / Glu↓Arg (Y)	$1.7\pm0.8*10^5$	$3\pm2*10^5$	Active and specific enzyme; electrostatic reversal of ion-pair specificity (N)
OmpT	Arg↓Arg / Tyr↓Arg (Y)	$1.7\pm0.8*10^5$	$8\pm3*10^4$	Active and selective enzyme (N)
OmpT	Arg↓Arg / Thr↓Arg (Y)	$1.7\pm0.8*10^5$	$2\pm1*10^4$	Active and selective enzyme (N)
OmpT	Arg↓Arg / Arg↓Val (Y)	$1.7\pm0.8*10^5$	$5\pm2*10^5$	~100-fold improvement in activity; efficient plasminogen activator. (Y)
OmpT	Arg↓Arg / Glu↓Ala (Y)	$1.7\pm0.8*10^5$	$1\pm0.9*10^6$	Highly active enzyme; neither P1 nor P1' are arginine (N)

Table 3.12: Specificity and selectivity of OmpT variants.

3.6 References:

1. Elsasser, S. & Finley, D. (2005) *Nat Cell Biol* **7,** 742-9.
2. Nadiri, A., Wolinski, M. K. & Saleh, M. (2006) *J Immunol* **177,** 4239-45.
3. Ohnishi, J., Ohnishi, E., Shibuya, H. & Takahashi, T. (2005) *Biochim Biophys Acta* **1751,** 95-109.
4. Rezaie, A. R. (2006) *Mini Rev Med Chem* **6,** 859-65.
5. Steinhoff, M., Buddenkotte, J., Shpacovitch, V., Rattenholl, A., Moormann, C., Vergnolle, N., Luger, T. A. & Hollenberg, M. D. (2005) *Endocr Rev* **26,** 1-43.
6. Mohamed, M. M. & Sloane, B. F. (2006) *Nat Rev Cancer* **6,** 764-75.
7. Tyndall, J. D., Nall, T. & Fairlie, D. P. (2005) *Chem Rev* **105,** 973-99.
8. Debela, M., Magdolen, V., Schechter, N., Valachova, M., Lottspeich, F., Craik, C. S., Choe, Y., Bode, W. & Goettig, P. (2006) *J Biol Chem* **281,** 25678-88.
9. Blanco, R., Carrasco, L. & Ventoso, I. (2003) *J Biol Chem* **278,** 1086-93.
10. Matsuura, T. & Yomo, T. (2006) *J Biosci Bioeng* **101,** 449-56.
11. McCarter, J. D., Stephens, D., Shoemaker, K., Rosenberg, S., Kirsch, J. F. & Georgiou, G. (2004) *J Bacteriol* **186,** 5919-25.
12. Kukkonen, M. & Korhonen, T. K. (2004) *Int J Med Microbiol* **294,** 7-14.
13. Walton, T. A. & Sousa, M. C. (2004) *Mol Cell* **15,** 367-74.
14. Vandeputte-Rutten, L., Kramer, R. A., Kroon, J., Dekker, N., Egmond, M. R. & Gros, P. (2001) *Embo J* **20,** 5033-9.

15. Varadarajan, N., Gam, J., Olsen, M. J., Georgiou, G. & Iverson, B. L. (2005) *Proc Natl Acad Sci U S A* **102,** 6855-60.

16. Grodberg, J., Lundrigan, M. D., Toledo, D. L., Mangel, W. F. & Dunn, J. J. (1988) *Nucleic Acids Res* **16,** 1209.

17. Hoover, D. M., Wu, Z., Tucker, K., Lu, W. & Lubkowski, J. (2003) *Antimicrob Agents Chemother* **47,** 2804-9.

18. Mirny, L. A. & Shakhnovich, E. I. (1999) *J Mol Biol* **291,** 177-96.

19. Graf, L., Craik, C. S., Patthy, A., Roczniak, S., Fletterick, R. J. & Rutter, W. J. (1987) *Biochemistry* **26,** 2616-23.

20. Cronin, C. N. & Kirsch, J. F. (1988) *Biochemistry* **27,** 4572-9.

21. Hwang, J. K. & Warshel, A. (1988) *Nature* **334,** 270-2.

22. Hedstrom, L., Szilagyi, L. & Rutter, W. J. (1992) *Science* **255,** 1249-53.

23. Olsen, M. J., Stephens, D., Griffiths, D., Daugherty, P., Georgiou, G. & Iverson, B. L. (2000) *Nat Biotechnol* **18,** 1071-4.

24. Gardell, S. J., Duong, L. T., Diehl, R. E., York, J. D., Hare, T. R., Register, R. B., Jacobs, J. W., Dixon, R. A. & Friedman, P. A. (1989) *J Biol Chem* **264,** 17947-52.

25. Lundrigan, M. D. & Webb, R. M. (1992) *FEMS Microbiol Lett* **76,** 51-6.

26. Stumpe, S., Schmid, R., Stephens, D. L., Georgiou, G. & Bakker, E. P. (1998) *J Bacteriol* **180,** 4002-6.

27. Peschel, A. & Sahl, H. G. (2006) *Nat Rev Microbiol* **4,** 529-36.

28. Hwang, B. Y., Varadarajan, N., Li, H., Rodriguez, S., Iverson, B. L. & Georgiou, G. (2006) *J Bacteriol*.

29. Okuno, K., Yabuta, M., Kawanishi, K., Ohsuye, K., Ooi, T. & Kinoshita, S. (2002) *Biosci Biotechnol Biochem* **66,** 127-34.

30. Aharoni, A., Gaidukov, L., Khersonsky, O., Mc, Q. G. S., Roodveldt, C. & Tawfik, D. S. (2005) *Nat Genet* **37,** 73-6.

31. Griswold, K. E., Aiyappan, N. S., Iverson, B. L. & Georgiou, G. (2006) *J Mol Biol* **364,** 400-410.

32. Collins, C. H., Leadbetter, J. R. & Arnold, F. H. (2006) *Nat Biotechnol* **24,** 708-12.

33. Sodeinde, O. A., Subrahmanyam, Y. V., Stark, K., Quan, T., Bao, Y. & Goguen, J. D. (1992) *Science* **258,** 1004-7.

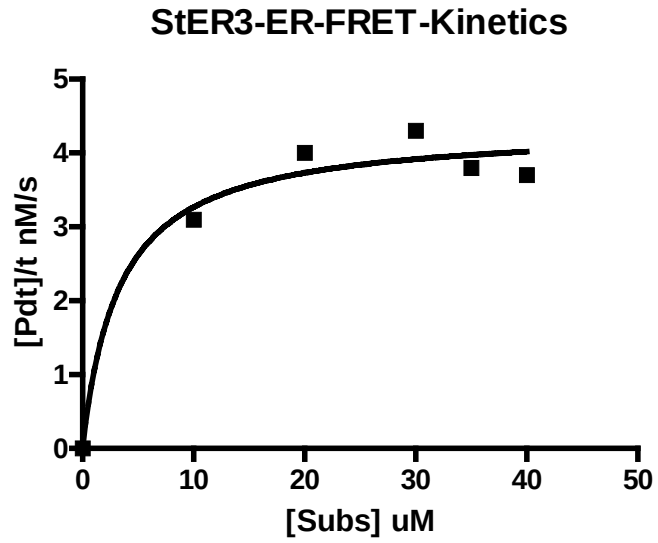
34. Kukkonen, M., Lahteenmaki, K., Suomalainen, M., Kalkkinen, N., Emody, L., Lang, H. & Korhonen, T. K. (2001) *Mol Microbiol* **40,** 1097-111.

35. Khersonsky, O., Roodveldt, C. & Tawfik, D. S. (2006) *Curr Opin Chem Biol* **10,** 498-508.

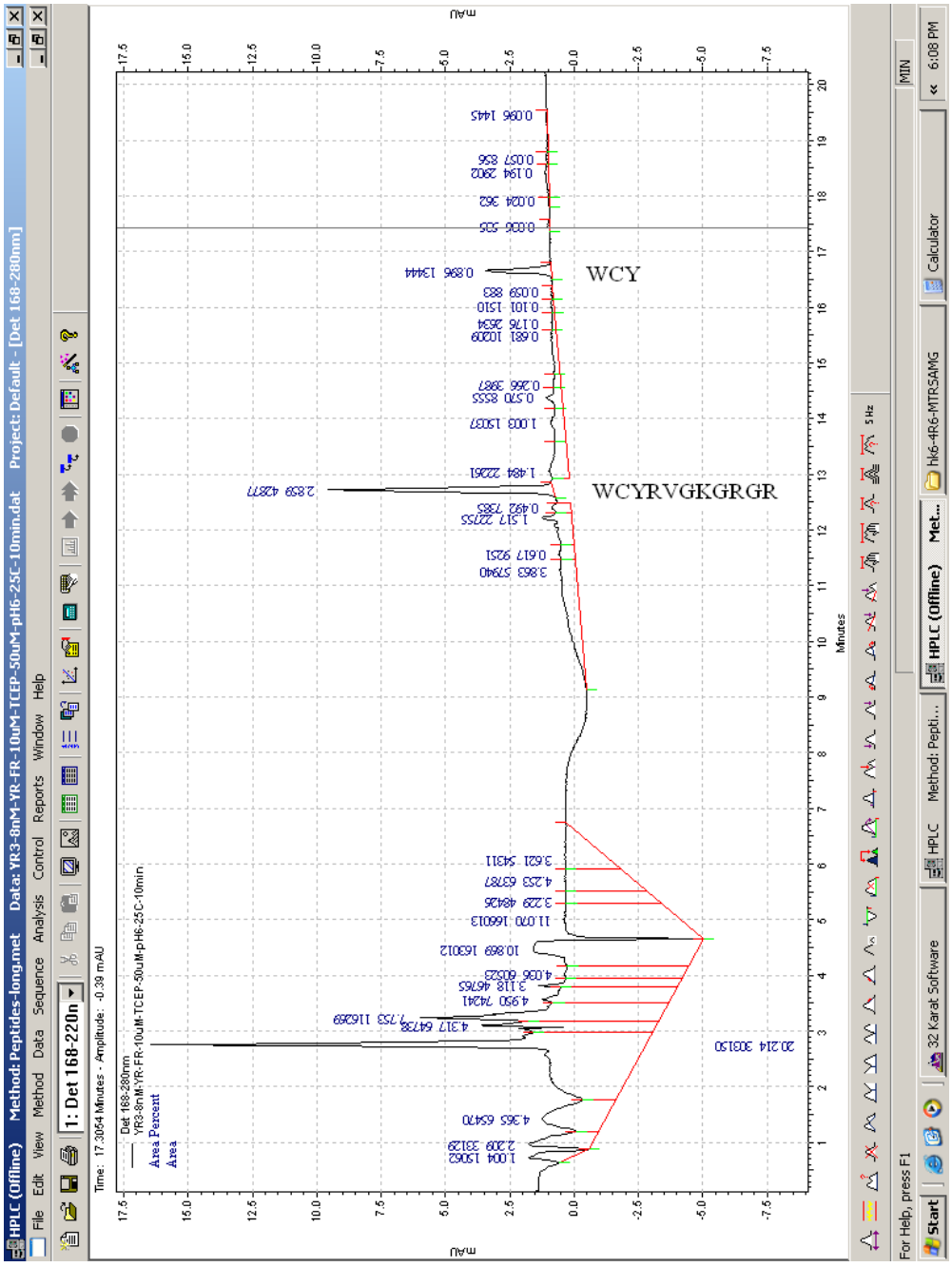
Appendix

A1. StER3-ER-FRET Kinetic data

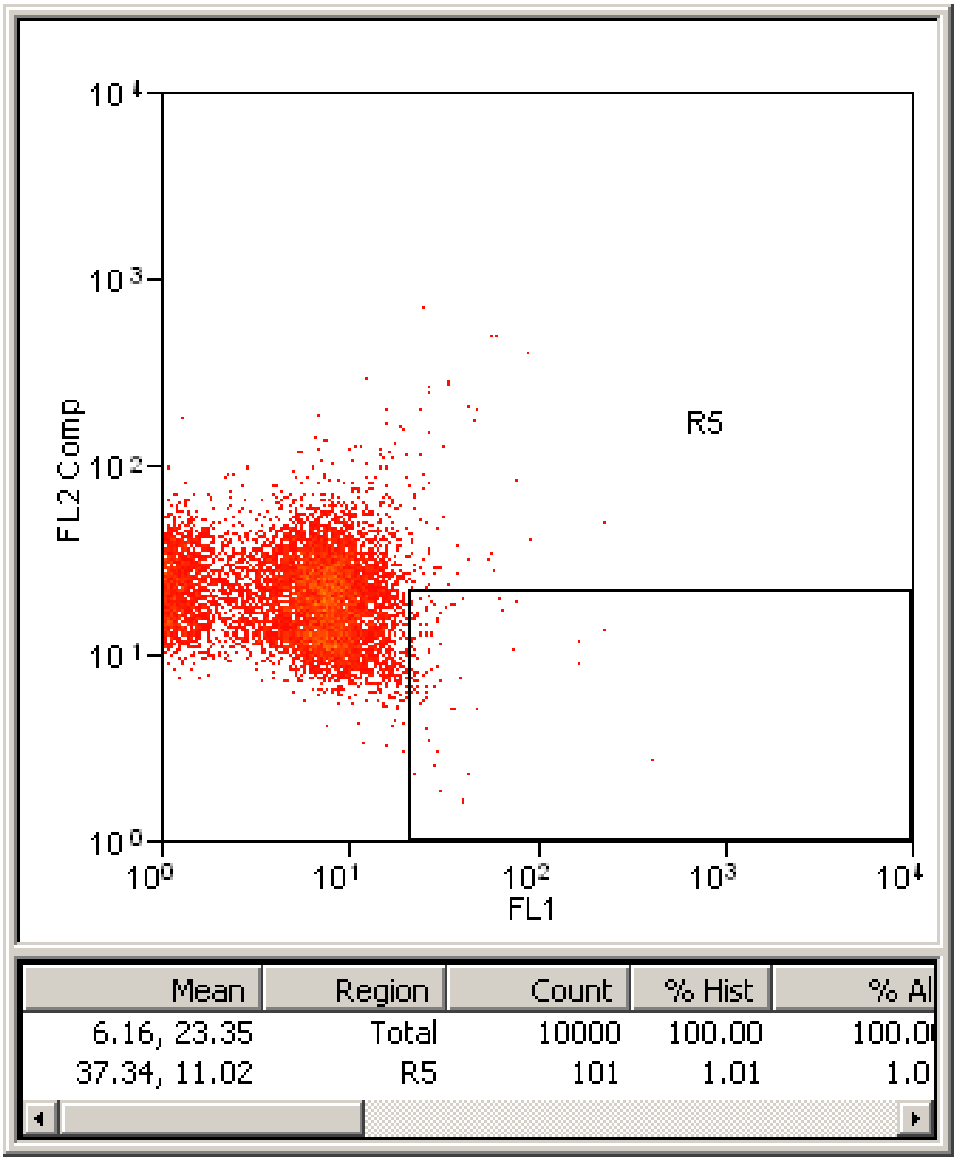
Substrate concentration (μM)	Product concentration/Time (nM/s)
0	0.0
10	3.1
20	4.0
30	4.3
35	3.8
40	3.7



A2. HPLC chromatogram (280nm) for the cleavage of the Tyr↓Arg substrate by 90YR3.



A3. Dot-plot of the Sat7 library labeled with 20 nM Thr↓Arg and 100 nM Arg↓Arg (10,000 events). R5 is the sort gate.



Mean	Region	Count	% Hist	% Al
6.16, 23.35	Total	10000	100.00	100.0
37.34, 11.02	R5	101	1.01	1.0

Vita

Navin Varadarajan was born in Mumbai, India on April 3, 1978, the son of T.R. Varadarajan and Vatsala Varadarajan. He grew up in the township of Chennai, India where he graduated from D.A.V. High School in 1995. He then entered RKM Vivekanada College (The University of Madras) in Chennai, India, where he graduated in 1998 with the degree of Bachelor of Sciences in Chemistry. In 1998, he joined the Indian Institute of Science, Bangalore, India and graduated with a Masters in Chemistry (M.S.) in 2001. In August 2001, he entered the Graduate School of The University of Texas at Austin to pursue his Doctorate in Chemistry.

Permanent address: D5 Browsnstone Apt, Mahalingapuram, Chennai, India 600034.

This dissertation was typed by the author.